A transferable coarse-grained model for peptides that display an environment driven conformational transition

by

Farhad Ramezanghorbani

A Thesis Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Computational Science and Engineering

Koç University

August, 2015

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a M.Sc. thesis by

Farhad Ramezanghorbani

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Asst. Prof. Dr. Mehmet Sayar

_____

Asst. Prof. Dr. Alkan Kabakçıoğlu

_____

Asst. Prof. Dr. Özge Şensoy

Date: _____

This thesis is dedicated to my family.

For their endless love, support and encouragement

ABSTRACT

Proteins structure and dynamics is strongly coupled to their functions. In order to study the dynamics and consequentially functions of these biomolecules, in addition to experimental and other theoretical approaches, molecular dynamics can be employed as a powerful tool. However, MD simulations are limited by the system size and simulation time. Processes like folding to native structures or aggregation take place on microsecond to second time scale, which is not accessible in MD simulations. In order to overcome the time and length scale obstacles coarse grained model has been employed as an alternative technique. When it comes to coarse-graining, transferability of coarse-grained (CG) model is one of the key challenges. Most of the state of the art CG models are parameterized to represent a single state point, and their validity upon changing the state point (different temperature, concentration, conformational change, or aggregation) is questionable and has to be carefully analyzed before utilizing.

In this thesis, we have developed a transferable coarse-grained model for proteins and peptides that can reproduce folding, aggregation and partitioning behaviour observed at hydrophobic/hydrophilic interfaces. To this aim a model synthetic peptide composed of leucine and lysine residues (with the sequence LKKLLKLLKKLLKL) has been used. LK peptide is designed to have built-in secondary amphiphilicity upon forming an α-helix. Recent studies have revealed that LK adopts several different secondary structures in solution (α-helix, β-hairpin, etc.), however none of them can dominate the conformational phase space. At a macroscopic (e.g air/water interface) or molecular (hydrophobic/hydrophilic interfaces created by neighboring molecules) interface, on the other hand, α-helix conformation dominates eliminating all other possible alternatives.

Our model is an implicit water model where the backbone of the peptide is represented by two beads and sidechains are represented by separate beads. The bonded interactions are tuned according to the Boltzmann distributions obtained from vacuum and tetramer simulations. The three main nonbonded interactions that play a key role in the model are: backbone hydrogen bonding, electrostatic interactions due to lysines, and leucine hydrophobic attraction in water and at the air/water interface. We have parameterized these interactions by using structural references derived from atomistic simulations. This model is verified by comparing the CG and atomistic potential of mean force curves for extension of the α-helix conformation in bulk water, adsorption and folding at the air/water interface, and finally aggregation of LK peptides in bulk water. Hence, our transferable coarse-grained model is capable of representing the behavior of the peptide in three different states with a single set of parameters.

ÖZETÇE

Proteinlerin yapıları ve dinamikleri, fonksiyonları ile güçlü bir şekilde bağlıdır. Bu biyomoleküllerin dinamiğini ve dolayısıyla fonksiyonlarını incelemek amacıyla, deneysel ve diğer teorik yaklaşımların yanında moleküler dinamik güçlü bir araç olarak kullanılabilir. Ancak, MD simülasyonları sistem boyutu ve simülasyon zamanı ile sınırlıdır. Doğal yapıya katlanma ya da kümelenme gibi süreçler mikrosaniye-saniye ölçeğinde gerçekleşmekte olup, MD simülasyonlarının bu ölçeklere erişimi mümkün değildir. Zaman ve boyut ölçeğindeki kısıtlamaları aşmak için düşük-çözünürlüklü modeller alternatif bir teknik olarak kullanılmaktadır. Düşük-çözünürlüklü modeller bahse konu olduğunda, modelin aktarılabilirliği en önemli konulardan biridir. Şu an var olan düşük-çözünürlüklü modellerin çoğunda sistemin faz uzayındaki tek bir nokta kullanılarak parametreler üretilir. Bu hal değiştirildiğinde (farklı bir sıcaklık, konsanstrasyon, konformasyon değişikliği ya da kümelenme) modelin geçerliliği sorgulanır hale gelmekte ve dolayısıyla kullanılmadan önce dikkatli bir şekilde test edilmesi gerekmektedir.

Bu tezde, biz katlanma, kümelenme ve ayrıca hidrofobik/hidrofilik arayüzlerde gözlemlenen ayrışma süreçlerini taklit edebilen, aktarılabilir düşük çözünürlüklü bir model hazırladık. Bu amaçla lösin ve lizinlerden meydana gelmiş (LKKLLKLLKKL-LKL dizilimine sahip) model bir sentetik peptit kullandık. Bu LK peptidi, bünyesinde alfa sarmalını oluşturduğu durumlarda açığa çıkacak ikincil amfifilik karakter barındırır şekilde tasarlanmıştır. Önceki çalışmalar LK peptidinin çözelti içerisinde farklı ikincil yapıları (alfa-sarmalı, beta-saçtokası ve benzerleri) da barındırdığını açığa çıkarmakla beraber, bu yapılardan hiçbirinin konformasyon faz uzayını domine edemediğine karar vermiştir. Öte yandan, makroskopik bir arayüzey (hava/su arayüzeyi) ya da moleküler bir arayüzeyin (komşu moleküller tarafından oluşturulan hidrofobik/hidro-

filik arayüzeyler) varlığında alfa-sarmalı konformasyonu tüm diğer olası seçenekleri eleyerek ortamı domine etmektedir.

Bizim modelimiz, örtülü su modeli olmakla beraber, peptit omurgası iki süperatomla temsil edilirken, yan gruplar ayrı süperatomlarla temsil edilmektedir. Bağlı etkileşimler vakum ve tetramer simülasyonlarından elde edilen Boltzmann dağılımlarına göre ayarlanmıştır. Modelde, bağlı olmayan etkileşimler arasında kilit rol oynayan üç ana etkileşim şunlardır: omurga hidrojen bağı, lizinlerden dolayı var olan elektrostatik etkileşimler ve suda ve hava/su arayüzeyinde etkili olan lösinler arası hidrofobik etkileşimler. Bu etkileimlerin parametreleri, atomistik simlasyonlar sonucunda elde edilen yapsal referanslar kullanlarak retilmitir. Bu modelin dorulamas iin dk-znrlkl modelle elde edilen ve alfa-sarmalnn su ierisinde ekilmesi, hava/su arayzeyine tutunmas ve katlanmas ve son olarak LK peptidinin su ierisinde kmelenmesi almalarndan elde edilen ortalama kuvvet erileri atomistik verilerle karlatrlmtr. Bu sayede, aktarlabilir dk znrlkl modelimiz tek bir parametre seti kullanarak farkl durumda peptit davrann baarl bir biimde temsil edebilmektedir.

ACKNOWLEDGMENTS

My sincere appreciations goes to my supervisor, **Dr. Mehmet Sayar** for providing me the chance to be a member of his group at Koç University. Things could not evolve without his trust and support. I am grateful to him for teaching me how to be a successful person and leader both in academia and real life as he is in his life.

I would like to thank my colleagues, **Cahit Dalgicdir** and **Beytullah Ozgur** for coming to my assistance with patience whenever I needed.

I am thankful for the financial support of Koç University and TÜBİTAK.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

MD	:	Molecular Dynamics
AA	:	All Atomistic
CG	:	Coarse-grained
REMD	:	Replica Exchange Molecular Dynamics
HREX	:	Hamiltonian Replica Exchange
PMF	:	Potential of Mean Force
PME	:	Particle Mesh Ewald
DoF	:	Degrees of Freedom
$\mathbf{E}_{hb}$	:	Epsilon value for Hydrogen Bonding potential
$\mathbf{E}_{w}$	:	Epsilon value for CG wall potential
$\mathbf{E}_{L}$	:	Epsilon value for leucine-leucine interaction potential
VS	:	Virtual Site

Chapter 1

INTRODUCTION

The properties of peptides and proteins are determined by structure and dynamic which are not unique for different environments. Peptides can be driven to different structures and consequentially display different functions due to change in PH, amphiphilicity, concentration, etc. of the corresponding environment. Amphiphilic nature of peptides causes different dynamics in the aqueous environment and at the molecular or macromolecular interfaces.

Folding and aggregation of amphiphilic peptides to create a hydrophobic core, adsorption to the hydrophobic/hydrophilic interfaces, and penetration through the membranes are examples of the dynamics at these environments. [6–8] In order to better understand the underlying pathways and mechanism for folding, aggregation, and partitioning the interface, molecular dynamic simulation is a valuable tool to analyze the atomistic behaviour and dynamic of systems. Classical molecular dynamics with all atom force field is limited to small systems and short simulation time scales. Typically folding and assembly can not be captured in a feasible time with all atom simulations. In order to overcome these barriers, coarse-grained models which have a lower resolutions are frequently employed. Coarse-graining allows usage of bigger size systems and longer simulation time by grouping atoms and treating them as single interaction sites (beads). [2]

Decreasing total number of particles by coarse-graining, results in reduction in degrees of freedom and speeds up the sampling. Simplified interaction potentials can be used for the coarse-grained interaction sites and energy barriers between different conformations is smoothed resulting in accelerated sampling of energy landscape. [9]

It is essential that interplay of folding, aggregation and partitioning at the interface is correctly represented in the CG model. Applicability of a CG model depends on the level of representability and transferability of the model. CG model being able to predict the properties of the system it is modeled after called representability and the model being able to predict different behaviour at different thermodynamic state points called transferability of model.

Transferability challenge discussed in several studies disclose the limitations associated with coarse-graining procedures [10–12]. Generally the nature of interaction potentials in a coarse-grained model has structure based considerations and is not purely energetic. In addition, discarding the entropic contributions of fast DoFs results in smoother energy surface compared to atomistic one (Figure 1.1).

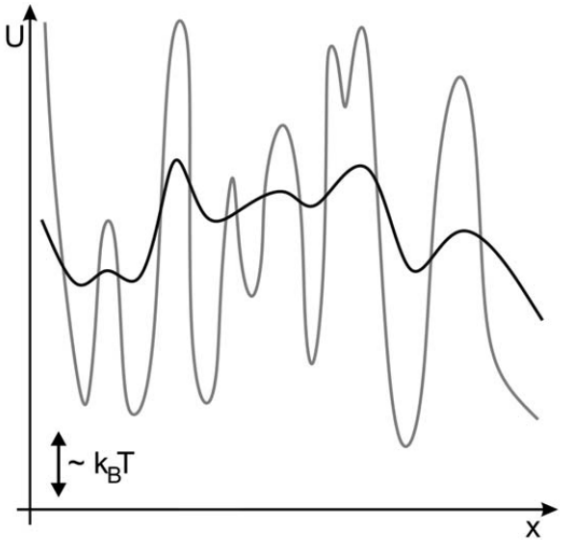

Figure 1.1: Shematic comparison of energy surfaces in atomistic system (grey line) and its CG representation (black line). Due to the decrease in DoFs, CG curve is smoother than the atomistic one, thus the CG model samples the configurational space faster. Achivement of transferability in CG model is limited due to the fact that equilibrium fluctuations that determine the thermodynamic state points are different in two energy surfaces [2].

The transferability of CG model is limited because the equilibrium states that corresponds to certain thermodynamic behaviour are different at two resolution levels. CG models and in general simulation models cannot be transferred to thermodynamic

conditions, concentrations, chemical compositions, etc., that deviate from the original state point where they had been parametrized.

Parametrization of bonded and non-bonded interactions for a coarse-grained model can be done using two overall approaches: bottom-up and top-down. In bottom-up approach, potentials can be derived to reproduce the quantities in higher resolution simulations (i.e. atomistic simulations) [13–15], and in the top-down approaches the development is done using macroscopic quantities observed through experiments [16] similar to parametrization of all atom force fields [17].

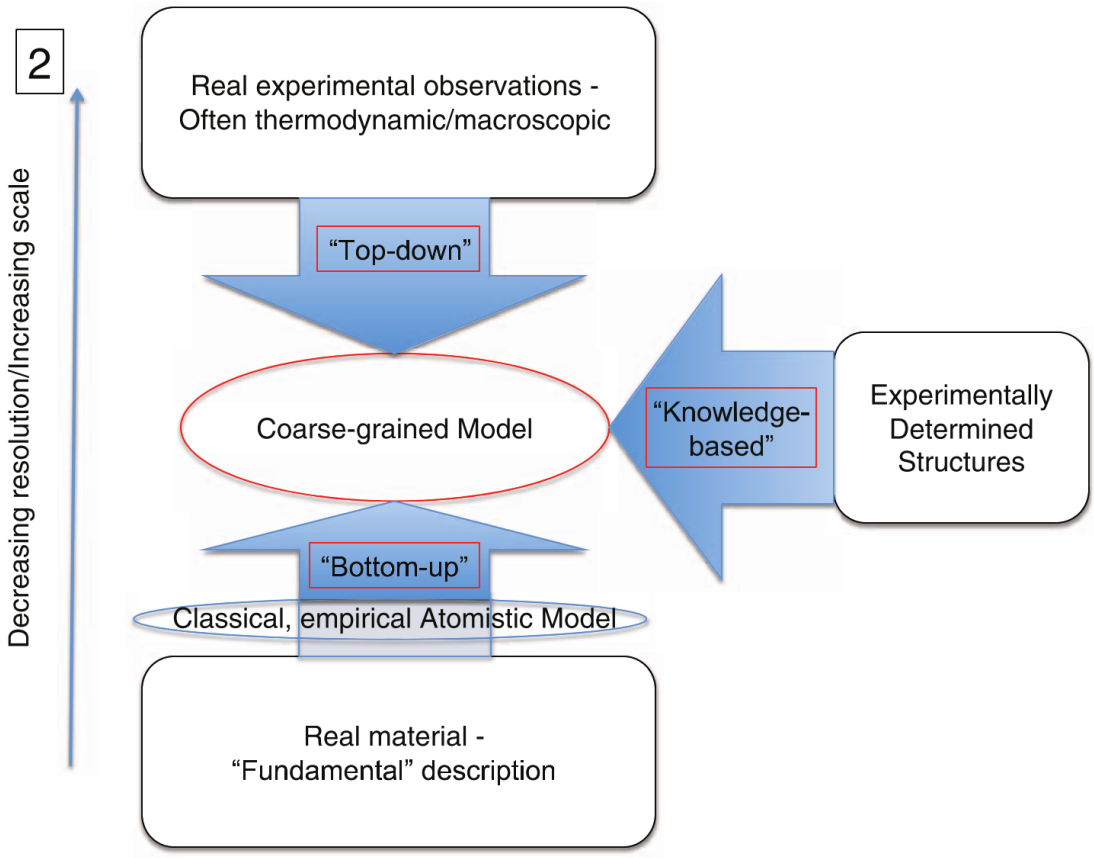

Figure 1.2: Schematic representation of different approaches for developping a coarse-grained model [3].

To model the flexibility of bonded DoFs, bond, angle and dihedral interactions can be generated using the direct Boltzmann inversion of the sampled bonded distributions.

[18, 19] Since bonds and angles can usually be considered *stiff* no strong coupling is presented with the nonbonded interactions, but this may not be the case for dihedrals. [2]

Non-bonded interactions that determine the important properties of a coarse-grained model can be constructed with two different methodological approaches: parametrized coarse-graining, and derived coarse-graining. In the first method target properties are calculated from atomistic simulations using systematic methods like iterative Boltzmann inversion [13], inverse Monte Carlo [20] or force-matching [21]. Transferability of the force-field is not guaranteed due to the structural based nature of methods in this case. In other words, these methods cannot predict non-target properties at the same or at a different thermodynamic state point that they are parametrized from. In the second method, direct atomistic interactions between superatoms which have a clear physical meaning are calculated. Solvation free energy [22] and pair potential of mean force [23] are examples of these methods. Force-field obtained from these energy based methods exhibit a good transferability. [15]

In this study we developed a transferable coarse-grained model for a synthetic lysine leucine repeating sequence (LK) which is capable of showing secondary amphiphilicity upon formation of α-helical structure. This 14 residues peptide adopts different secondary structures in bulk water and at the hydrophobic/hydrophilic interfaces. LKα14 that has an intrinsically disordered nature at low concentration in solution, displays a conformational transition to α-helix upon adsorption to the interface and upon aggregation. [1, 24, 25] It is found that the interplay of hydrophobic interactions and hydrogen bonding induces such a conformational transition. We aim to build a CG model for LK which can successfully reproduce the conformational behaviour of the molecule in bulk water, at the air water interface and in aggregates. The ability of LK to fold to α-helix, partially helical and β-sheet secondary structures and form different aggregate sizes make this molecule an ideal model to extend the coarse-graining study to more complicated models.

The purpose of this thesis is to develop a CG force-field that displays both representability and transferability of LKα14 in different environments. Chapter 2 explains the methods and tools used in this study and introduces the mapping scheme proposed for the CG model. In Chapter 3, the procedure for constructing bonded interactions is reviewed and modifications done to the initial parameters are discussed. In Chapter 4, 5, and 6 nonbonded interactions have been introduced to the model. Definition of hydrogen bonding with virtual sites representation, together with structure based Hamiltonian replica exchange, and PMF based checks are included in Chapter 4. The focus of Chapter 5 is interface-related hydrophobic interactions. Construction of CG interface based on density profile has been done, and interactions tuned using adsorption free energy of LK to the interface. Chapter 6 closes the nonbonded interaction parameterization by introducing hydrophobic interactions that contribute to the emergence of a hydrophobic core upon aggregation. As a conclusion, Chapter 7 presents overall check for the CG model in the presence of all parameters and summarizes our findings.

Chapter 2

METHODS

2.1 Mapping Scheme

The sequence of LK peptide is $Ace(LKKLLKL)_2Nme$ where L stands for Leucine, and K for Lysine sidechains. To develop the coarse-grained force field based on buttom-up approach several all atomistics simulations have been used. Initial mapping scheme for LKα14 that was suggested by Cahit Dalgicdir [1] had 5 different super atoms. This primary model included two backbone beads named **CA** and **CN**, where CA contains single C_α atom and the remaining main chain C, O, N, H atoms are represented by CN bead. In this model Leucine sidechains are represented with one **L** bead, and Lysine sidechains with two **KC** and **KN** beads where the former contains the $CH_2CH_2CH_2$ and the latter includes the CH_2NH_3 atoms. All of the superatoms are placed at the center of mass of the atom groups they represent. (Table 2.1)

beads	**mass**	**atoms**
CN	43.025	$CONH$
CA	13.019	C_α
L	57.116	$C_\beta C_\gamma C_{\sigma 1} C_{\sigma 2}$
KC	42.081	$C_\beta C_\gamma C_{\sigma 1}$
KN	42.081	$C_\eta\ NH_3$
H	0	$VirtualSite3out$

Table 2.1: Definition of the beads and their masses in the CG model of LKα14. H is a virtual interaction site constructed based on "3out" virtual interaction site type implemented in Gromacs. Its location has been calculated based on neighboring CAs and CN superatoms

The atomistic LKα14 is mapped to the CG model using the mapping scheme presented in detail by Table 2.1. Here in Figure 2.1 we have CG and AA extended structures mapped on top of each other. Hydrophobic Leucines are colored in red and hydrophilic Lysine sidechains in blue. In this figure, virtual sites for hydrogen are not presented.

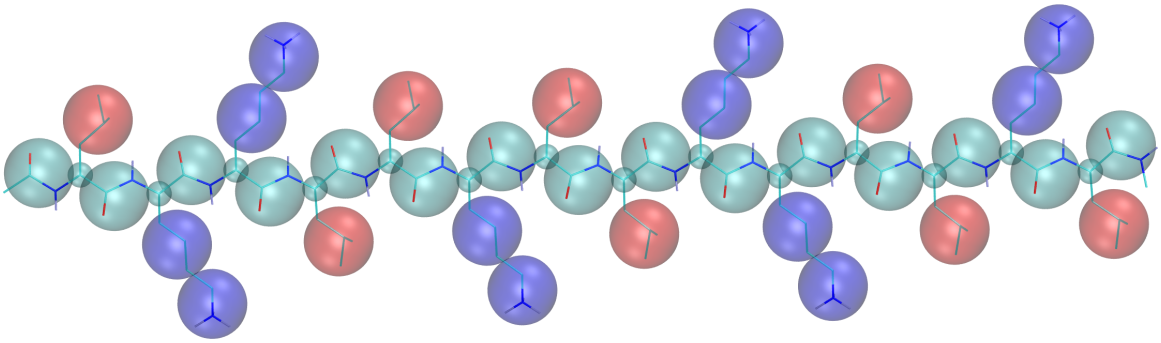

Figure 2.1: The initial CG mapping scheme for LKα14. In this extended conformation coarse-grained beads are on top of atomistic counterparts and their radii were tweaked to contain the corresponding atoms. The colouring of beads are as follows: **L** are red, **KC** and **KN** are blue , backbone beads (**CA** and **CN**) are cyan. [1].

Later (as will be discussed during the course of the thesis), we have added extra modifications to our primary model to include virtual sites representation of hydrogen bonds. The final model includes hydrogen beads (**H**) with zero mass where their positions are calculated based on three other neighboring beads (CA,CN,CA). To be able to define the first and the last H beads we need two additional C_α at the two ends of peptide on top of the NME and ACE alpha carbons. The final model has 16xCA, 15xCN, 6xL, 8xKN and 8xKC beads, and 15 virtual sites corresponding to each hydrogen beads. Figure 2.2 shows the atomistic LK and coarse-grained model in helical conformation with the presence hydrogen beads.

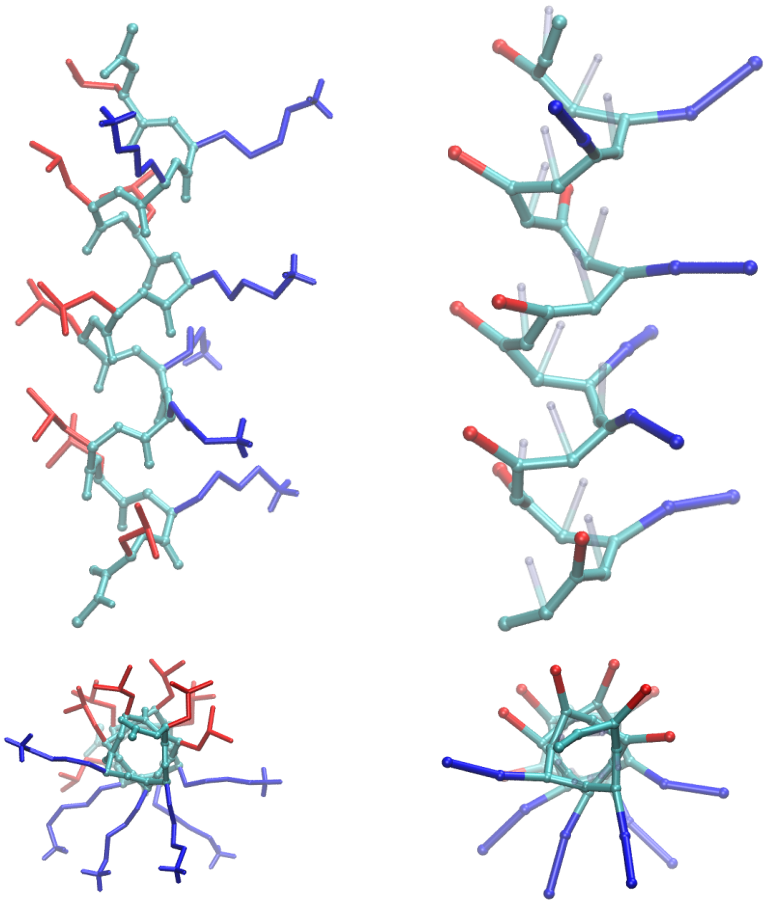

Figure 2.2: Atomistic LKα14 (left) representation and the final CG mapping scheme (right) both side and top view of the helix are given. Virtual interaction sites for hydrogen bonded to CN beads are also shown. The colouring scheme of beads are as follows: **L** are red, **KC** and **KN** are blue , backbone beads (**CA** and **CN**) are cyan, and **H** are grey.

2.2 Simulations and Analysis

VOTCA package [26] has been used to map the all atomistic references onto the CG model representation of the molecule. KN beads are protonated and chloride ions inserted to the system to maintain charge neutrality. CG molecular dynamic simulations were conducted using GROMACS version 4.5.6 [27] with the leap-frog stochastic dynamics integrator and a 2 fs time-step. We performed all the MD simulations in canonical ensemble (NVT) to avoid large volume fluctuations. Temperature was set to

298° K using velocity-rescaling algorithm [28] with coupling time constant of 1 ps. Van Der Waals (VDW) interactions were truncated at 1.0 nm. Long range electrostatic calculations were taken into account using particle mesh Ewald (PME) [29] with a coulomb cut-off of 2.0 nm and fourier spacing of 0.24 nm to increase the efficiency of the simulations. To be consistent with all atomistic references, for single LK in bulk simulations we used a cubic box of length 4.5 nm for dimer simulation to have the same volume as the dodecahedron box in the atomistic case. In order to introduce the Gromacs hard wall for representing the interface we increased the z dimension of the simulation box to mimic the same volume in atomistic interface. Energy minimization using steepest descent algorithm were applied to all coarse-grained systems.

Umbrella sampling [30] method was used for potential of mean force (PMF) [31] calculations in four different cases: pulling single LK from hard wall (interface), pulling peptide from a dimer aggregate, pulling peptide from tetramer aggregate and pulling two ends of LK in helix conformation. A series of initial conformations has been generated for umbrella windows and we performed an MD simulation in NVT ensemble for 200 ns using each conformation. Weighted histogram analysis (WHAM) [32] using g_wham toolkit [33] of GROMACS package was then applied to obtain the PMF curve. The convergence for PMF was checked by block analysis.

For pulling peptide from the interface we pulled the center of mass of the backbone beads away from a dummy particle which was freezed at the wall location and have no interaction with the wall. Pulling was done in the z direction perpendicular to the interface with the pulling rate of 0.005 nm/ps using a force constant of 1000 $kJ/(mol.nm^2)$. Pulling LK from a dimer aggregate has been done using the center of mass of the backbone beads for pull and reference groups. A pulling rate of 0.005 nm/ps in three dimensions with the same force constant as interface pulling was applied. For pulling two ends of an α-helix we used two CA beads at two ends of peptide (CA0 and CA15) with the pulling rate of 0.01 nm/ps and the force constant of 1000 $kJ/(mol.nm^2)$ in three dimensional geometry.

Hamiltonian replica exchange [34] implemented in GROMACS 4.6.1 [35] patched

with the PLUMED plugin [36] version 2.2.0 has been used to enhance sampling of conformational space of CG model. 16 replicas were setup for HREX simulations, which have different Hamiltonians $H(p, q)$ instead of different temperature. HREX allows us to use independent topology files for each replica, thus we have used 16 different scaling factors correspond to the hydrogen bonding potential's depth varying from 14 kJ/mol to 17.75 kJ/mol. HREX simulation was conducted for 2 μs, in implicit solvent with NVT ensemble. A conformation exchange between neighboring replias was attempted every 1000 steps and the average exchange probability was determined to be 0.84.

All visualizations in this thesis were produced by VMD [37].

Chapter 3

BONDED INTERACTIONS

In order to obtain bonded interaction potentials for the CG model a single peptide was simulated in implicit solvent with all the nonbonded interactions excluded. In this case peptide behaves randomly and adopts a random coil structure. Then the resulting trajectory has been mapped using our CG mapping scheme and from this output trajectory, probability distributions for bonds, angles and dihedrals has been calculated. The bonded distributions obtained from this simulation have the contribution of bonded interaction minus the coupling with nonbonded interactions. From now on we will call this all atomistic simulation as *vacuum-excl.* simulation.

A comparison between *vacuum-excl* and tetramer AA simulations for all bonded distributions has been done. [1] Bond stretching distributions yield close curves, and for angle distributions, weak correlation with non-bonded interactions has been observed. The differences between tetramer and *vacuum-excl* angle distributions is considered to be the result of many body interactions in the aggregated state. Since some bonded degrees of freedom like bonds and angles can usually be considered *stiff*, no strong coupling is presented with the weaker nonbonded interactions. But this might not be the case for dihedrals. [2] The backbone dihedral distributions reveal the difference between the helical conformation in tetramer and the extended one in vacuum. Consequently, since the dominant conformation in bulk water is known to be helix, half-helix or β-hairpin, and also in order to avoid strong coupling of bonded and nonbonded interactions we constructed the bonds and angles from tetramer reference and dihedral potentials using *vacuum-excl.*

3.1 Bonds stretching interactions

Bond stretching distributions from tetramer AA reference simulation have been Boltzmann inverted and resulting potentials have been fitted to a harmonic function. Spring constants and equilibrium distances for these harmonic potentials are provided in Table 3.1. Based on the period of these harmonic oscillations, proper time step for CG simulations has been chosen to be 2 fs. [5, 38]

Bond	r_0 **(nm)**	K_r $(kJ/(mol.nm^2))$
CN-CA	0.2112	95210.0
CA-CN	0.1960	157766.0
CA-L	0.2630	62476.3
CA-KC	0.2590	85700.8
KC-KN	0.3194	72640.4

Table 3.1: Harmonic potential parameters for bond stretching in CG model. Equilibrium bond length (r_0) and spring constant (K_r) for each bond have been given. [1]

3.2 Angle bending interactions

Tetramer angle distributions has been used to construct the tabulated angle potentials for our CG model. Two backbone angles (CA-CN-CA and CN-CA-CN), and three side chain angles (CN-CA-L, CN-CA-KC and CA-KC-KN) had been suggested in the initial model. Later through the further modifications two additional side chain angles have been added which we will refer as *reverse angles*. Reverse angles L-CA-CN and KC-CA-CN has been constructed to limit the area spanned by side chain beads. All the angles distributions has been Boltzmann inverted to obtain the corresponding tabulated potentials shown in Figure 3.1.

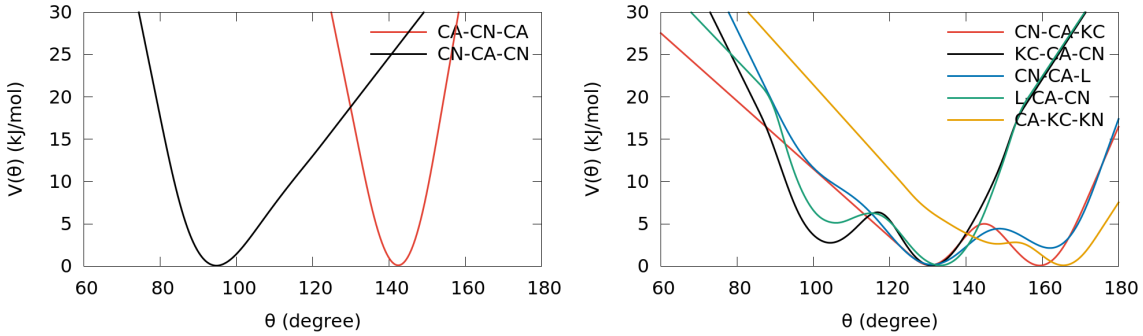

Figure 3.1: Tabulated angle potentials employed in the CG model for backbone angles (left) and side chain angles (right). Angles interaction potentials are Boltzmann inversion of their corresponding distributions in tetramer atomistic simulation where LK adopts an ideal helical conformation.

3.3 Torsional angle interactions

Figure 3.2 is the comparison between tetramer backbone dihedral distributions and implicit water (*vacuum excl.*) ones. As these probability distributions suggest, one of the major differences between an extended structure in implicit solvent and a helical conformation in the tetramer is backbone dihedral distribution which reveals this conformational switch.

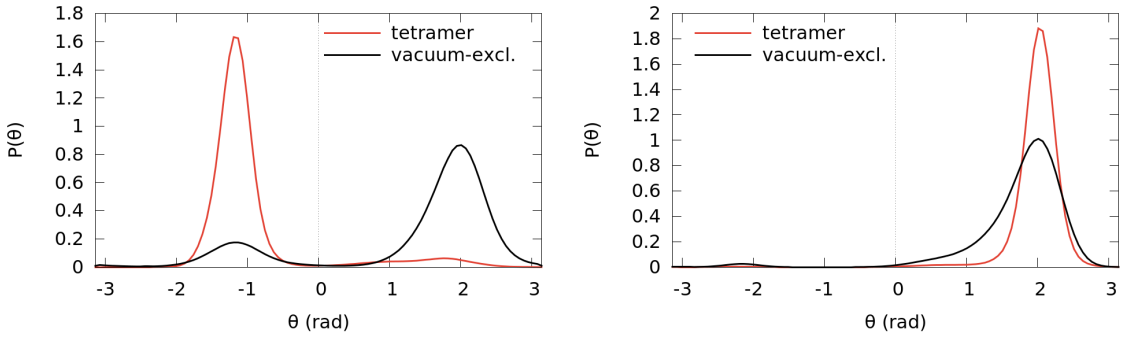

Figure 3.2: Backbone dihedral angle distributions for CN-CA-CN-CA (Left) and CA-CN-CA-CN (Right). Backbone dihedrals are a good signature of peptide's helicity. For CN-CA-CN-CA, left side of the distribution corresponds to α-helix and right side to an extended structure. In CA-CN-CA-CN, the height of probability distribution reveals the difference between secondary structure.

CN-CA-CN-CA dihedral angle is where we can clearly see this secondary structure transition in bonded parameters. We will use this distribution to track the behaviour of system while tuning the hydrogen bonding interactions.

For our CG model, dihedral distributions from *vacuum-excl.* has been Boltzmann inverted to construct the corresponding tabulated potentials. In addition to regular backbone dihedrals, two improper dihedral angle has been constructed for L and K side chains. [1] Figure 3.3 illustrates the tabulated interaction potentials used for torsional angles in the CG model.

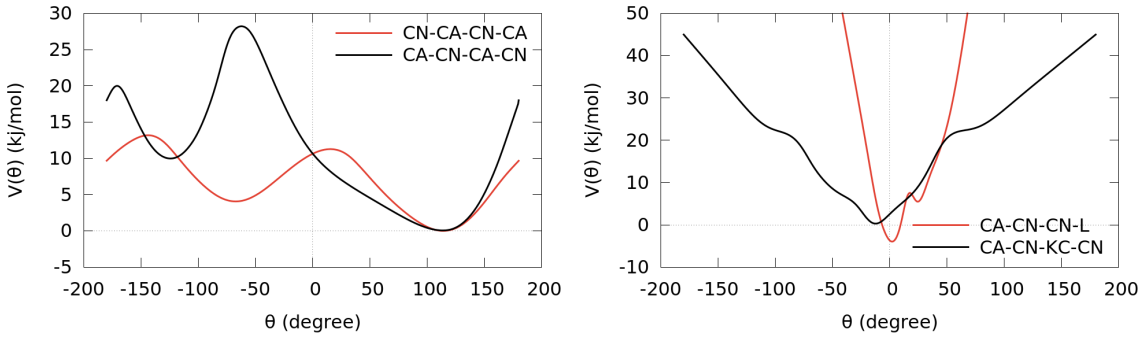

Figure 3.3: Torsional angles tabulated potentials for backbone dihedrals (Left) and side chain improper dihedrals (Right). Torsional angles interaction potentials are Boltzmann inversion of their corresponding distributions in atomistic *vacuum-excl.* simulation where LK adopts extended conformation. [1]

When the CG model is simulated in implicit solvent with these bonded interaction potentials, the distribution of CA-CN-CA-L dihedral angle creates a minor peak at 120 degree, while such a dihedral angle was not discovered in all atomistic simulations. A dummy dihedral provided has been constructed to prevent sampling of this unwanted region by CA-CN-CA-L dihedral distribution. [1] Later in 3.4, introducing reverse angles, this unwanted region has not been sampled, thus dummy potential will not be presented in our final model.

3.4 Further modification of bonded interactions

As it is expected, by using only bonded interactions of previous section we are not able to see helix formation in implicit water. It is known that three diving forces

that determine the conformational state of LK are hydrogen bonding, hydrophobic interactions and electrostatics. [1]

The simplest approach would be to use a pair potential among CN beads to represent hydrogen bonds. This specific pair potential has been chosen to be Boltzmann inversion of $CN_i - CN_{i+4}$ distance distribution in atomistic tetramer (blue curve in Figure 3.4). [1] Adding this hydrogen bonding potential we aim to reproduce random behaviour in bulk and helical conformation at the CG interface.

The bonded interactions presented in our model have been used together with a nonbonded pair potential for hydrogen bonding in an implicit water simulation. To mimic the hydrophobic/hydrophilic interface at the CG scale we have used simple wall definition in GROMACS. This CG wall favors L beads as the CG counterpart of the atomistic hydrophobic Leucines and disfavors Lysine (KN, KC) side chains. Later we will change the wall representation by adding more details based on the density profiles of different beads at AA interface simulation.

The backbone dihedral angles (CN-CA-CN-CA, and CA-CN-CA-CN) and the hydrogen bond distance ($CN_i - CN_{i+4}$) contain the difference regarding the transition between a helical conformation in tetramer (or at hydrophobic/hydrophilic interfaces) and random structure adopted in bulk water (or REMD). Trying to mimic atomistic REMD reference distributions for backbone dihedrals and $CN_i - CN_{i+4}$ distance we found the proper scaling factor for pair potential. However our model failed to show the stable helical structure in the wall simulation with the same scaling parameter. Even starting with an α-helix at the wall, LK unfolds throughout the simulation.

It has been shown that in the absence of pair potential for hydrogen bonding, interaction potential of wall with L beads plays no part in the $CN_i - CN_{i+4}$ distance distribution, and therefore the helical structure. [1] Considering the pair potential to be the only parameter that induces a stable helical conformation in this model, I started modifying this tabulated potential to make it more distance specific.

Boltzmann inversion of the 6_{th} $CN_i - CN_{i+4}$ distance distribution (which is located in the middle of peptide) in tetramer AA reference simulations has been used. Extrap-

olating the repulsive part and applying a cutoff at 0.7 nm, the blue curve in Figure 3.4 has been constructed. In comparison to the previous pair potential (red curve in Figure 3.4) the new potential is a better representation of the distance dependent nature of hydrogen bonding.

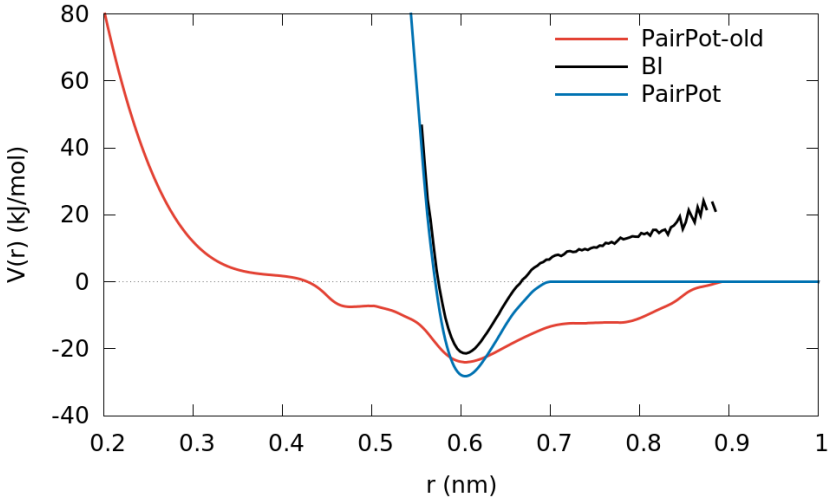

Figure 3.4: Pair interaction potential applied for $CN_i - CN_{i+4}$ pairs in the old model (red), together with the modified version of it (blue) that is more distance specific. To construct this new pair potential, $CN_i - CN_{i+4}$ distance distribution has been Boltzmann inverted (black) and truncated at 0.7 nm. This pair potential is not presented in the final CG model of LK, instead a generic definition of hydrogen bonding will be applied.

With the new tabulated pair potential for hydrogen bonding, *vacuum-excl.* dihedral potentials, angles and bonds interaction potentials from tetramer, we performed MD simulation for LK model in implicit solvent. Even though this strong pair potential (-28 kJ/mol) keeps LK in its helical conformation, there is a contribution to *extended conformation* part of backbone dihedral distribution (right side of CN-CA-CN-CA). After investigating this issue by checking each dihedral distribution, we discovered that this non-helical contribution is valid even for backbone dihedrals located in the middle of LK. (Figure 3.5)

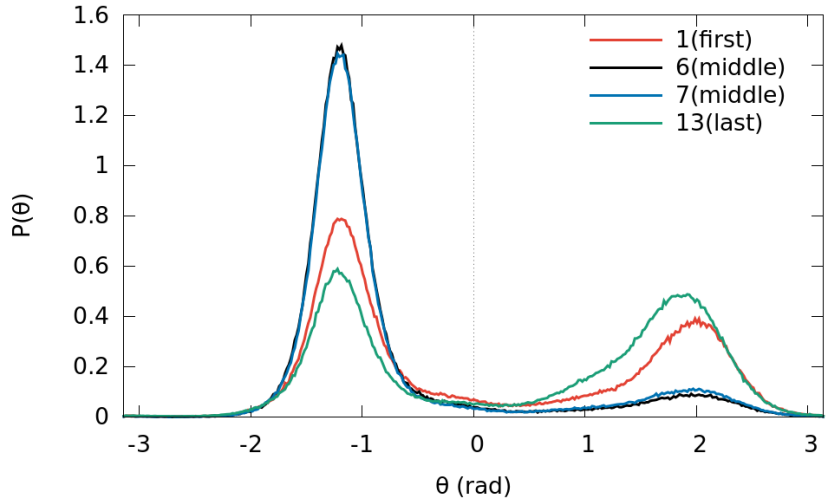

Figure 3.5: Contribution of CN-CA-CN-CA dihedrals located in the middle of LK model to non-helical conformations.

In addition to the backbone dihedral distributions, jumps to non-helical distances in time-line data of $CN_i - CN_{i+4}$ has been observed, suggesting that there may be strong repulsions in our new pair potential which are responsible for these big jumps in distance, as shown in Figure 3.6.

To check if there are irregularities in each of the bonded distributions, besides fitting a harmonic function to pair potential well to find appropriate time-step, a harmonic bonded potential for $CN_i - CN_{i+4}$ pairs has been used. Even in these cases we could observe jumps in time-line data of hydrogen bonding pairs.

Turning the nonbonded and bonded interaction potentials off, we ran NVE simulations in which we turned on angles and dihedrals one by one to check if the energy of the system diverges or not. This enables us to see which tabulated potential(s) is responsible for the energy divergence.

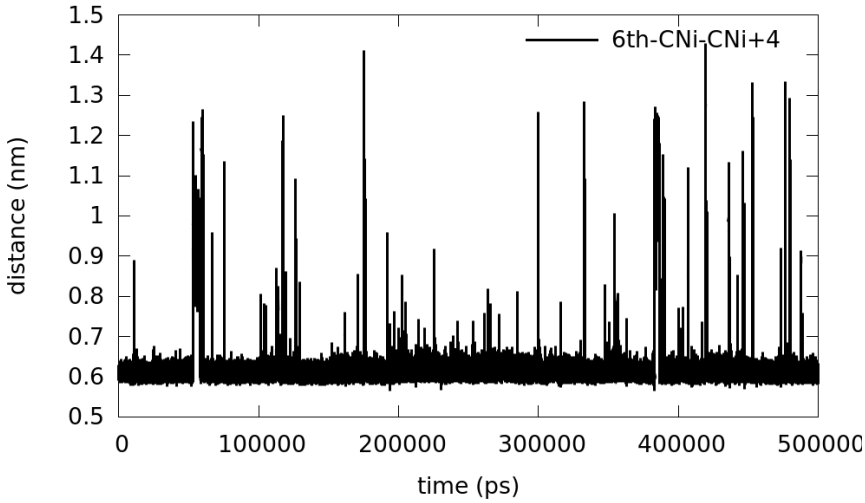

Figure 3.6: $CN_i - CN_{i+4}$ distance for a pair located in the middle of LK during the course of 500 ns simulation at the interface. Equilibrium distance for hydrogen bonding pairs ($CN_i - CN_{i+4}$) is 0.6 nm. Fast distance jumps are a consequence of the sharp repulsion in the pair potential.

Even by turning off all the dihedrals and non-bonded potentials we had energy divergence and system blowed up. So we started with smoothing out the forces for angle tabulated potentials and restarted the NVE simulation without dihedral potentials. In this case system did not crash in NVE simulation in the absence of dihedrals.

Figure 3.7 shows the total energy of the system during a 100 ns NVE simulations in the presence of different combination of dihedral interactions. Energy of the system is constant only when improper dihedrals (CA-CN-CN-L, and CA-CN-KC-CN) and dummy dihedral (CA-CN-CA-L) are off.

3.4.1 Reverse angles

By closer look at the atoms involved in these dihedrals we found out that for impropers and dummy dihedrals there are angles involved which can take the value of 180 in their distributions (CN-CA-L and CN-CA-KC) and consequently cause problem in dihedral energy calculations. In this case mentioned dihedrals rely on just three non-planar

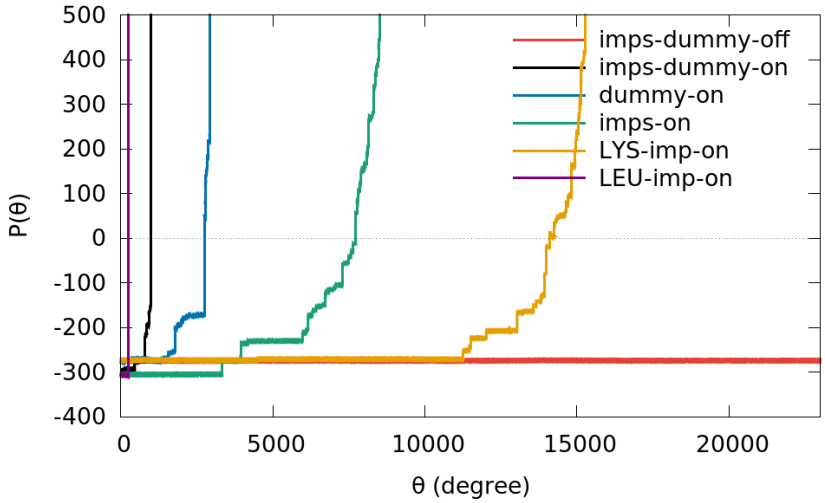

Figure 3.7: Total energy of system in an NVE simulation with the presence of different combination of dihedral interactions. In the absence of dummy dihedral and impropers (side chain), energy of system does not diverge.

atoms to define plane normal.

To fix this, *reverse angles* (L-CA-CN and KC-CA-CN) have been added to our angle interaction potentials that restrict the area spanned by Leucine and Lysine side chains. These new angle potentials constructed based on the corresponding atomistic distributions of tetramer simulation. For Leucine side chain we need to use both CN-A-L and L-CA-CN interaction potentials and for Lysine, CN-CA-KC and KC-CA-CN. (Figure 3.8)

By using both angle potentials for side chains and removing the dummy dihedral we observed that the total energy of the system was conserved. Adding these angles which are constructed based on tetramer AA, backbone and improper dihedrals obtained from *vacuum-excl.* and pair potential obtained from 6_{th} $CN_i - CN_{i+4}$ pair in tetramer AA, NVT simulation in implicit solvent has been performed to check the over-restraining of side chains due to the additional potentials.

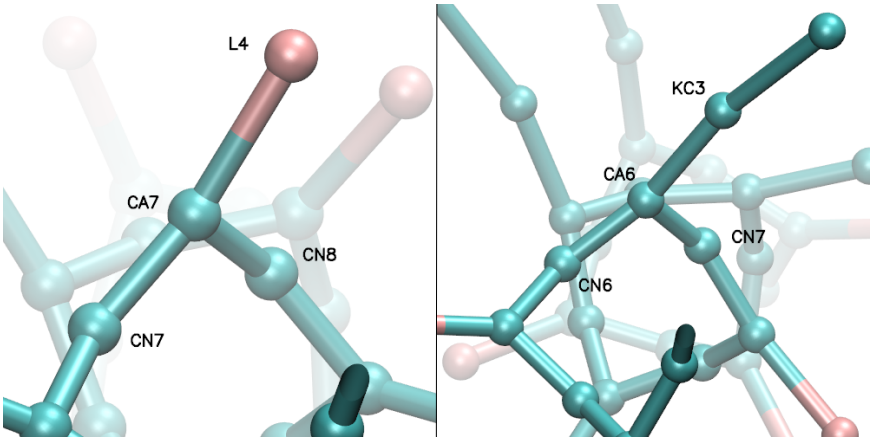

Figure 3.8: Snapshots of LK CG in helical conformation where individual Leucine and Lysine side chains are labeled as an example to illustrate new angle potentials added to the system. In this Figure L4-CA7-CN7 and KC3-CA6-CN6 are examples of new side chain angles.

Figure 3.9 shows the comparison for side chain angle distributions in our CG model vs AA references. In all angles and their reversed cases we see that CG result is overshooting the tetramer distribution for the corresponding angle. One approach is to decrease these side chain angle potentials by a factor of 0.5 instead of using the direct Boltzmann inverted forms.

Using scaled (by 0.5) side chain potentials we could not find clear results that scaled angle will solve the over restraining issue. In the original side chain angles (CN-CA-L and CN-CA-KC) it could solve the overshooting problem, but in the reverse cases (L-CA-CN and KC-CA-CN) it could not reproduce the corresponding tetramer AA reference distributions. (green curve in Figure 3.9)

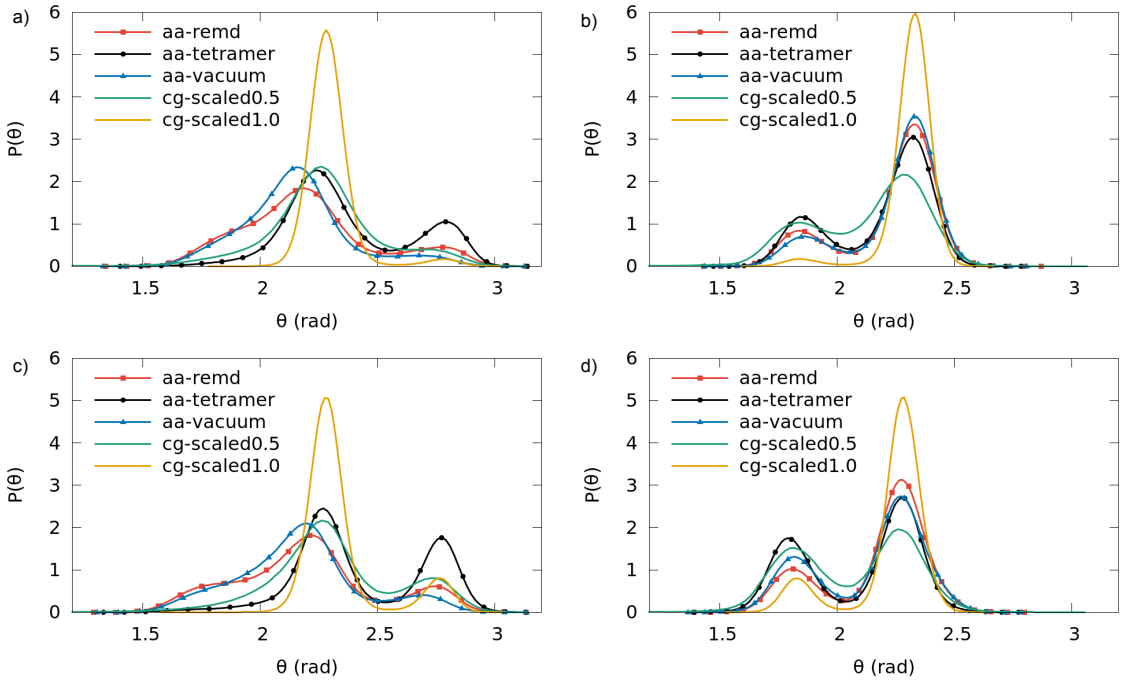

Figure 3.9: Side chain angles probability distributions, for CN-CA-L (a), L-CA-CN (b), CN-CA-KC (c) and KC-CA-CN (d), comparing with atomistic REMD, *vacuum-excl.* and tetramer reference distributions.

Necessity of improper dihedrals for side chain has been checked throughout a 100 ns simulation of LK in implicit water. Comparison between corresponding improper dihedral distributions in CG and AA references reveals the fact that even with reverse angle potentials, side chains cannot sample the correct region. Distributions for side chain dihedrals has been distorted in the absence of improper dihedrals. Figure 3.10 shows this comparison with and without the presence of improper dihedrals.

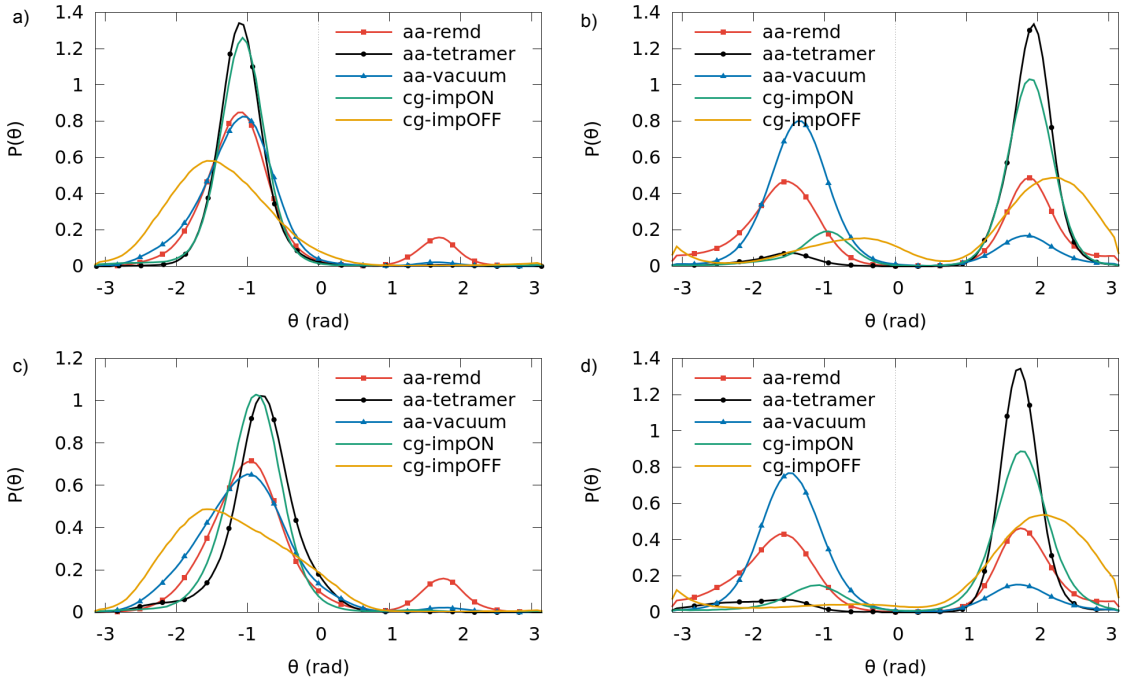

Figure 3.10: Side chain improper dihedrals probability distributions, for CA-CN-CA-L (a), L-CA-CN-CA (b), CA-CN-CA-KC (c) and KC-CA-CN-CA (d), comparing with atomistic REMD, *vacuum-excl.* and tetramer reference distributions.

Clearly, we could not reproduce all the atomistic distributions due to the decrease in DOF forced by coarse-graining scheme. Bonded interactions are constructed so that LK is not biased towards any conformation. Due to the selection of backbone dihedral interactions it has the similar behaviour to the *vacuum excl.* simulation and we will see LK model is capable of making different conformation observed in bulk in the presence of a generic hydrogen bonding interaction. In the next chapters we will introduce and tune three non-bonded interactions which trigger the conformational changes in LKα14.

Chapter 4

NON-BONDED INTERACTIONS: HYDROGEN BONDING

As expected, using bonded interaction potentials obtained in previous chapter, the CG model fails to reproduce the $CN_i - CN_{i+4}$ distance distribution of the α-helix structure in atomistic tetramer. Therefore one needs additional interactions to describe hydrogen bonding and amphiphilicity of peptide. In the following chapters we will introduce three types of interactions to our CG model: Hydrogen bonding, Hydrophobic attractions, and Electrostatic interactions. We will use structure based and thermodynamics based approaches for each type to construct the interaction parameters. Our aim is to observe the transition between a random structure in bulk and helical conformation in the presence of interface or aggregates. Since transition between random structure and α-helix is visible in the distance distribution of $CN_i - CN_{i+4}$ and the CN-CA-CN-CA dihedral distribution, we will use them as a guide to reproduce the REMD results in bulk and atomistic tetramer at interface and in the presence of other peptides.

The simplest form of the potential for hydrogen bonding interactions is a specific pair potential for the $CN_i - CN_{i+4}$ pair, which has been already explained in the previous chapter. Later we see that even though the hydrogen bonding is happening between the i_{th} and $(i+4)_{th}$ residues in nature, choosing the current mapping scheme for our model changes the corresponding indices to the i_{th} and $(i+3)_{th}$ CN beads. Later we will switch to $CN_i - CN_{i+3}$ pairs to employ the hydrogen bonding potential.

With the modified pair potential and initial wall interactions explained in the previous chapter (3.4) our helical LK is not stable at the CG interface and unfolds

quickly. Starting with an extended structure, our peptide adsorbed to the interface but again failed to switch to a helical conformation. Moreover, it has been shown that in the presence of CG interface, this pair potential is the only parameter that affects helix formation [1]. In other words when we do not have pair potential as a driving force for helicity, wall interactions only induce partitioning of the amphiphilic residues. Using an attractive LJ potential for Leucines, a repulsive one for other beads, and employing the wall definition of Gromacs also fails to reproduce the correct partitioning at the interface. Leucines are smashed to the interface resulting in an unfolded structure. (Figure 4.1) This inability of the wall to induce a change in secondary structure suggests that we need to use a more detailed combination of these non-bonded interactions.

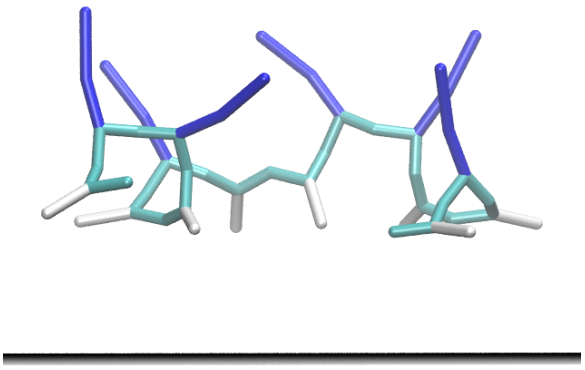

Figure 4.1: Snapshot of CG LK at the Gromacs hard wall. The wall is shown with a black line. Leucine residues which are colored in white, are smashed to the interface and lose the correct partitioning.

Other drawbacks of using simple pair potential between specific pairs is that LK is unable to represent β-hairpin conformation where different donors and acceptors can involve in hydrogen bonding. In addition to that, in helical conformation, we have strange hydrogen bond formation for the first and the last pairs which satisfies the distance cutoff for pair potential but lacks a direction dependency. (Figure 4.2) These hydrogen bonds happen in the cost of the transition of the backbone dihedral. In other words the first and the last residues flip and sample the non-helix region of the

backbone dihedral, but at the same time satisfy the CN-CN equilibrium distance for hydrogen bonding.

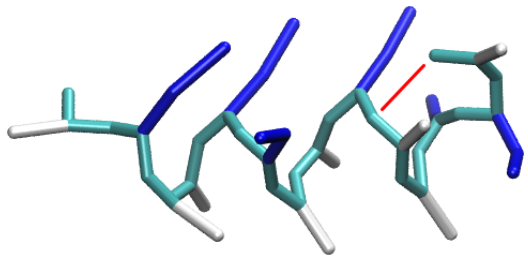

Figure 4.2: Snapshot of CG LK in helical conformation with the pair potential interaction for hydrogen bonding. On the right end of LK last residue flipped and still satisfies the hydrogen bonding distance. The red line shows the unphysical hydrogen bonding for the terminal residue.

Our aim is to use a generic potential for hydrogen bonding interactions that shows both the distance and direction dependent nature of hydrogen bonding. The interaction depends on the distance and orientation of the amide group and carbonyl group. Atomistic amid group is composed of a nitrogen and a hydrogen whereas the carbonyl one has a carbon and an oxygen. The direction dependent nature of the hydrogen bond prefers C=O...H-N atoms to be aligned while making hydrogen bond. Geometrical criterion for hydrogen bonding which is used by Gromacs suggests distances less than 0.35 nm between donor and acceptor (N-C) and HNC angles less than 30 degrees.

4.1 Virtual sites representation

All suggested CG models by literature for hydrogen bonding potential have explicit representation for all the backbone atoms. [4, 39–42] In these models, positions of hydrogen and oxygen atoms were calculated using the local geometry of backbone beads. Then a Lennard-Jones potential combined with an angular term has been used to describe the hydrogen bonding. (Figure 4.3)

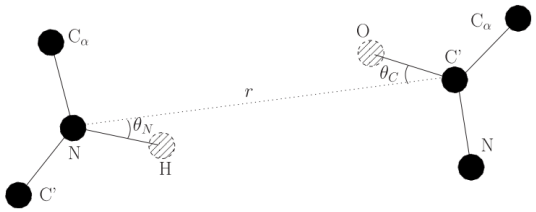

Figure 4.3: Schematic representation of hydrogen bonding interaction using virtual sites, taken from [4]. The backbone in Bereau's model has been mapped to three beads and shaded beads are not explicitly presented in the model. The position of these beads are inferred from their bonded neighbors.

Since our CG model represents backbone with only two beads (CA and CN), by using the backbone geometry we cannot define the H and O locations. But instead we have used three consecutive backbone beads CA-CN-CA plus one other backbone bead (receptor) CN_{i-3} to construct a virtual interaction site. The position of the virtual interaction site or dummy particle in GROMACS can be found in different ways using 2, 3 and 4 other atoms. Since the position of VS is a function of the coordinates in actual beads, the force acting on the virtual site must be redistributed over the particles with mass in a consistent way. In this way the total force is conserved. [5]

Among 6 different types of virtual site construction (*2, 3, 3fd, 3fad, 3out, 4fdn*), we have used 3out in which the position of VS is a non-linear combination of three other beads, and out of their plane.

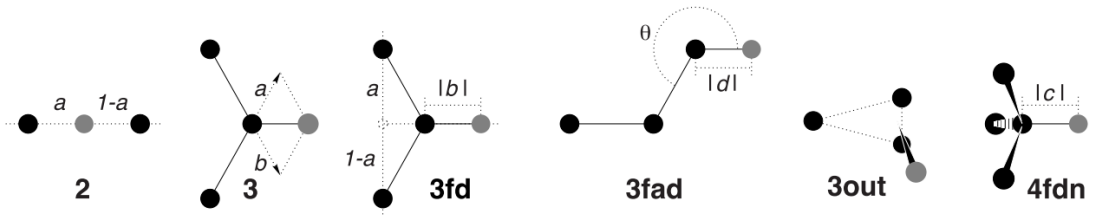

Figure 4.4: Different types of virtual interaction site representations implemented in Gromacs. The naming is based on number of atoms involved in construction of VS. [5]

In this representation we defined VS so that it is located in the middle of $CN_i - CN_{i+3}$ distance (r_s) in the helical conformation. The VS location has to be out of the plane constructed by CA_{i+2}, CN_{i+3}, and CA_{i+3} (j,i, and k in Figure 4.5).

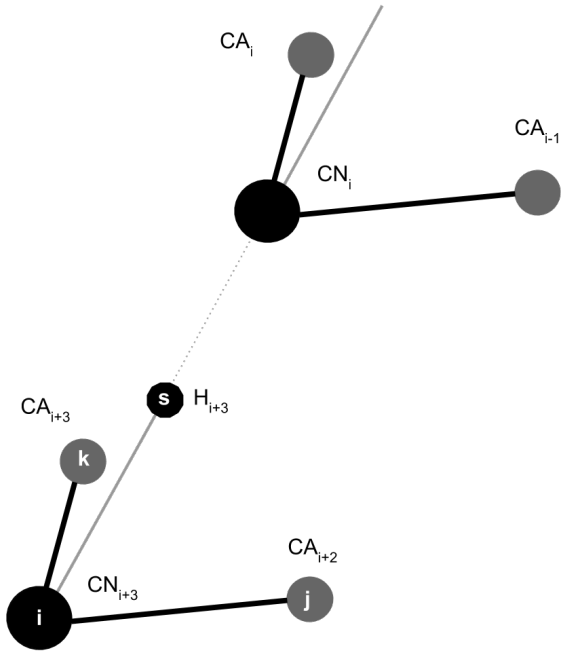

Figure 4.5: Schematic representation of hydrogen virtual site in LK CG model. VS location has been calculated based on three other beads (i,j,k) so that it locates in the middle of $CN_i - CN_{i+3}$ distance which is out of the plane constructed by (i,j,k).

Virtual site positions have been found from $CN_i - CN_{i+3}$ distance distribution in an ideal helix in tetramer atomistic reference. Using a set of one attractive potential (Eq. 4.1) for CN-H, and two repulsive potentials (Eq. 4.2) for CN-CN and H-CA, we

mimicked the directionality of hydrogen bonding. Additional exclusions have been applied between CN-CN and H-CN pairs with i and i+2 indices to prevent unwanted hydrogen bonding. To employ exclusions for virtual sites in Gromacs we need to add a fake harmonic bond with a force constant of zero for virtual site ($CN_i - H_i$).

Attractive CN-H potential is a modified Lennard-Jones 12-6 potential (Eq. 4.1) with a cosine term to be able to die out faster in longer distances compare to normal 12-6 LJ potential. The minimum (r_m) has been assigned to 0.235 nm based on $\frac{1}{2}(CN_i - CN_{i+3})$ distance distribution from atomistic tetramer and the cutoff (r_c) is set to 0.3 nm so that it prevents making two hydrogen bonds with CN_{i+3} and CN_{i+4} at the same time. Epsilon value will be tuned based on structural based and PMF based approaches.

$$\begin{cases} U = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad \text{where} \quad \sigma = r_m \times \sqrt[6]{\frac{1}{2}} & r \leq r_m \\ U = -\epsilon\left[\cos\left(\frac{\pi(r-r_m)}{2(r_c-r_m)}\right)\right]^2 & r_m < r < r_c \\ U = 0 & r_c \leq r \end{cases} \tag{4.1}$$

Even in the presence of CN-CN repulsion, CN_{i+4} can get close to the hydrogen bonding distance by distorting the helix geometry to benefit from this irrelevant hydrogen bonding. To keep CN_{i+4} away from hydrogen bead, and preserve helical conformation we need to add another generic repulsive potential for CA-H pairs. Equation 4.2 illustrates the repulsive potential used for the mentioned pairs. Initial ϵ value for these potentials has been set to 100 kJ/mol. For CN-CN interactions r_m has been set to 0.480 nm and for CA-H to 0.338 nm. Later, these r_m values will be re-tuned after decreasing the repulsive ϵ to 10 kJ/mol.

$$\begin{cases} U = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon \quad \text{where} \quad \sigma = r_m \times \sqrt[6]{\frac{1}{2}} & r \leq r_m \\ U = 0 & r_m < r \end{cases} \tag{4.2}$$

Now that we have a generic definition for hydrogen bonding potential, tuning the corresponding parameters (E_{hb}, r_m, r_c) our aim is to reproduce different conformations in implicit solvent. Different approaches has been employed to find the proper values for the parameters of the hydrogen bonding potentials which will be discussed in the following.

4.2 Structural based approach

LKα14 adopts different secondary structures in bulk (α-helix, β-hairpin, half-helix and random-coil) and is able to switch between different conformations. Since atomistic REMD has been used to sample the conformational space in bulk, and we aim to mimic the bulk behaviour with CG model in implicit solvent, CG probability distributions will be compared with REMD-AA reference distributions. Backbone dihedrals and hydrogen bonding distance are the target distributions which clarify the random coil to helix transition.

In this section we are tuning the hydrogen bond interaction by using the backbone dihedral angle distribution. We chose the backbone dihedral because that is a good signature of helical conformation of LK molecule. For CN-CA-CN-CA, if LK is helical, probability distributions switches to the left and when it is non-helical it switches to the right.

At this point there is no charge in the system and electrostatic has not been taken into account, there is also no hydrophobic interactions for Leucine side chains. We will recall the ϵ value for CN-H hydrogen bonding potential as E_{hb} from now on. With different E_{hb} values, Single LK has been simulated in implicit solvent for 1 μs and the probability distributions have been compared with REMD-AA ones in Figure 4.6.

Resulting distributions in Figure 4.6 shows an energy barrier between the ϵ values

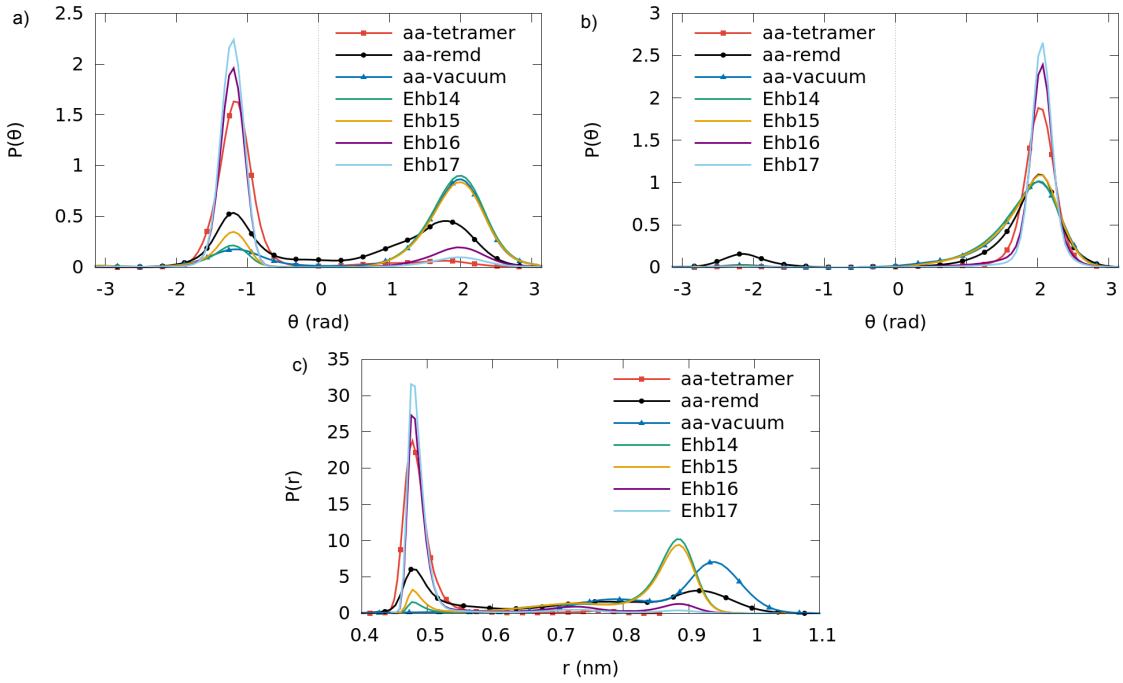

Figure 4.6: Backbone dihedral distributions, CN-CA-CN-CA (a), CA-CN-CA-CN (b) and hydrogen bonding distance distributions, ($CN_i - CN_{i+3}$) (c) compared with AA reference distributions. Generic hydrogen bonding potential with different ϵ values for CN-H potential (E_{hb}) has been used for LK model in implicit solvent. The bimodal behaviour of CG distributions disclose an energy barrier between the helical and extended structures.

(E_{hb}) 15 and 16 kJ/mol. Lack of sampling the conformational space results in time-dependency of these distributions. This bimodal behaviour is also observable for the decimal values between 15 and 16 (15.1, 15.2, ...) but not in an ordered fashion. Since using this misleading time dependent distributions makes it questionable to find the proper value for E_{hb}, we took an additional criterion to check the helix - random coil transition.

Similar to the random behaviour in bulk reported by Dalgicdir et al. [1], LK model in implicit solvent is expected to show frequent folding-unfolding events. Thus folding-unfolding behaviour throughout the simulations has been monitored for different E_{hb} values varying from 15.1 to 15.9. The minimum E_{hb} value in which LK shows frequent folding-unfolding in a $2\mu s$ simulation in implicit solvent is 15.6 kJ/mol. With other

values, LK either unfolds quickly and remains extended for the rest of simulation or does not completely unfold and stays helical in a $2\mu s$ simulation. Figure 4.7 is the time-line data of $(CN_i - CN_{i+3})$ distances while E_{hb} is 15.6 kJ/mol. Starting from a helical conformation LK model can unfold to a completely extended conformation then folds back to a helical structure. LK model is able to find and make hydrogen bonds using this generic hydrogen bonding definition with E_{hb} 15.6 kJ/mol.

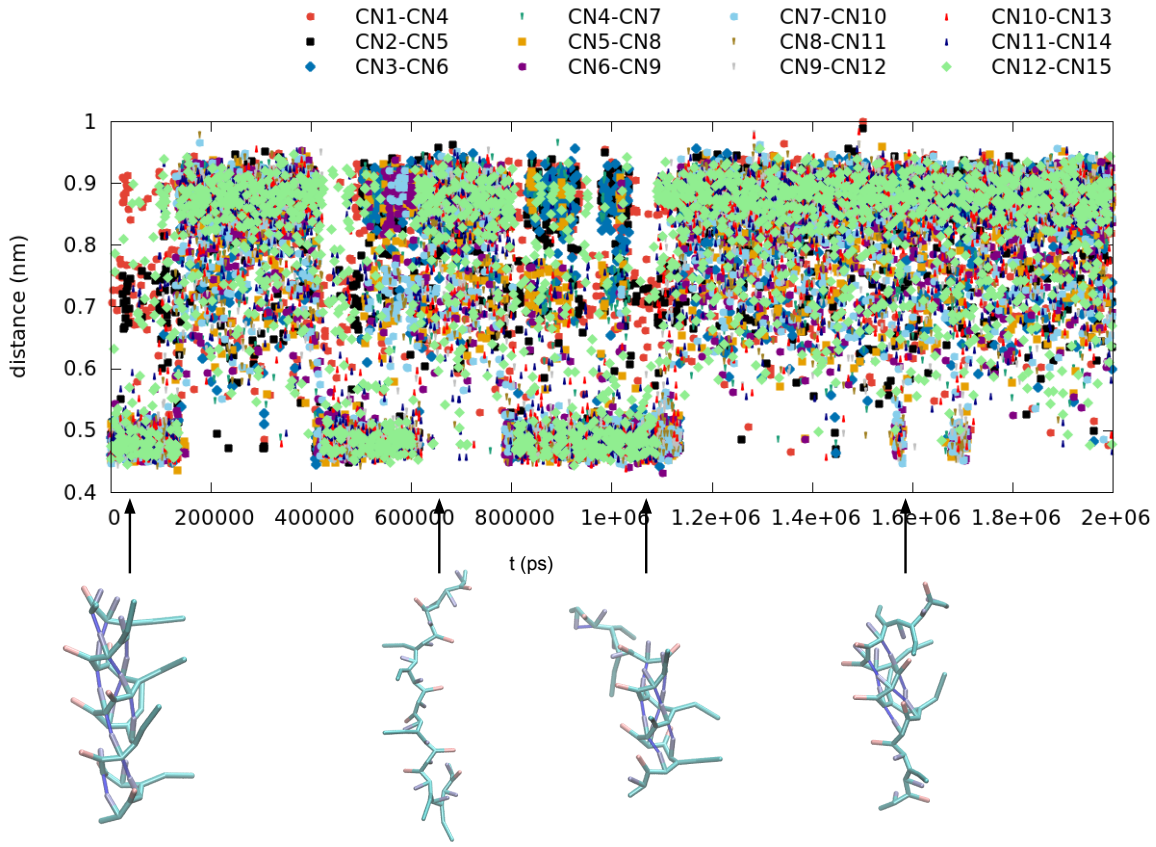

Figure 4.7: Hydrogen bonding distances (for $CN_i - CN_{i+3}$ pairs) versus time, for a single LK model simulated in implicit solvent for $2\mu s$. Starting from an ideal α-helix, LK unfolds completely and folds back again to an ideal helix. Snapshots of LK, adopting different conformations have been attached to their corresponding times in the simulation.

Extreme fluctuations observed in the time-line data of hydrogen bonding distance together with the inability of LK model to fold back and make hydrogen bonds while starting from the extended conformation at the interface, suggests that it is still

hard for hydrogens in our model to find the corresponding acceptor beads (CN) and make hydrogen bonds. Using sharp repulsive potentials (H-CA, CN-CN) with ϵ 100 kJ/mol, makes it hard for hydrogen to find the equilibrium distance with CN and passes through all these energy barriers. To overcome these sharp energy barriers, we smoothed the repulsive part of all tabulated potentials and tuned their r_m them so that we can mimic the correct helical conformation. For CA-H new r_m is 0.358 nm and for CN-CN, 0.495 nm.

For CN-CN and H-CA, we decreased the ϵ value of generic repulsion from 100 to 10 kJ/mol. For CN-H attractive potential we are using a modified Lennard-Jones potential which has cosine term to let the attractive part die out faster in longer distances. Cutoff has been increased to 0.31 nm so that H can parse a larger area for CN beads. Cutoff (r_c) is short enough to prevent making hydrogen bond with both CN_{i+3} and CN_{i+4} at the same time. For the repulsive part of CN-H, before r_m, we have used ϵ value of 10 kJ/mol to make repulsion smoother. Then this repulsive part shifted down in the y direction to make a continuous potential curve. Figure 4.8 shows the smoothed set of potentials for hydrogen bonding.

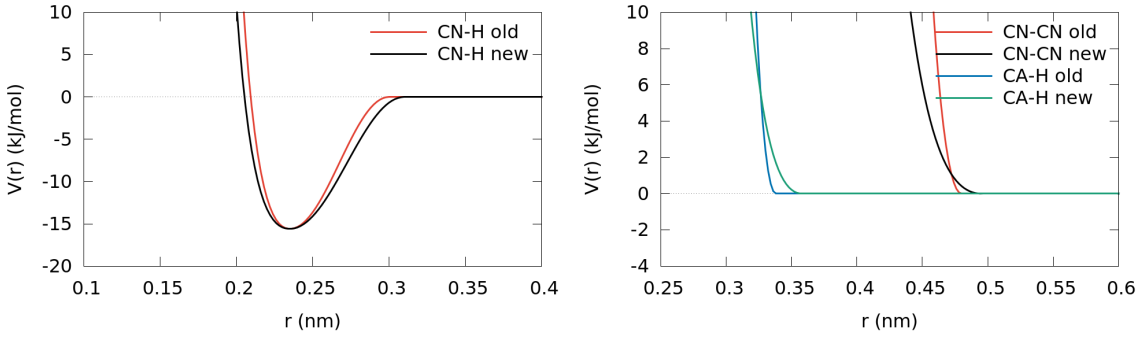

Figure 4.8: Modified hydrogen bonding interaction potentials. Epsilon value for the repulsive parts of these potentials has been decreased, and r_m has been shifted to mimic the correct distributions for backbone dihedrals and hydrogen bonding distances in helical conformation. The cutoff (r_m) for CN-H potential has been increased to allow the hydrogen beads to cover a broader area searching for the CN beads.

Using these smoothed repulsions, in the presence of CG interface which will be defined in the next section, extended structure adsorbed to the interface and gradually

folds back to its helical conformation in a $1\mu s$ simulation. (Figure 4.9)

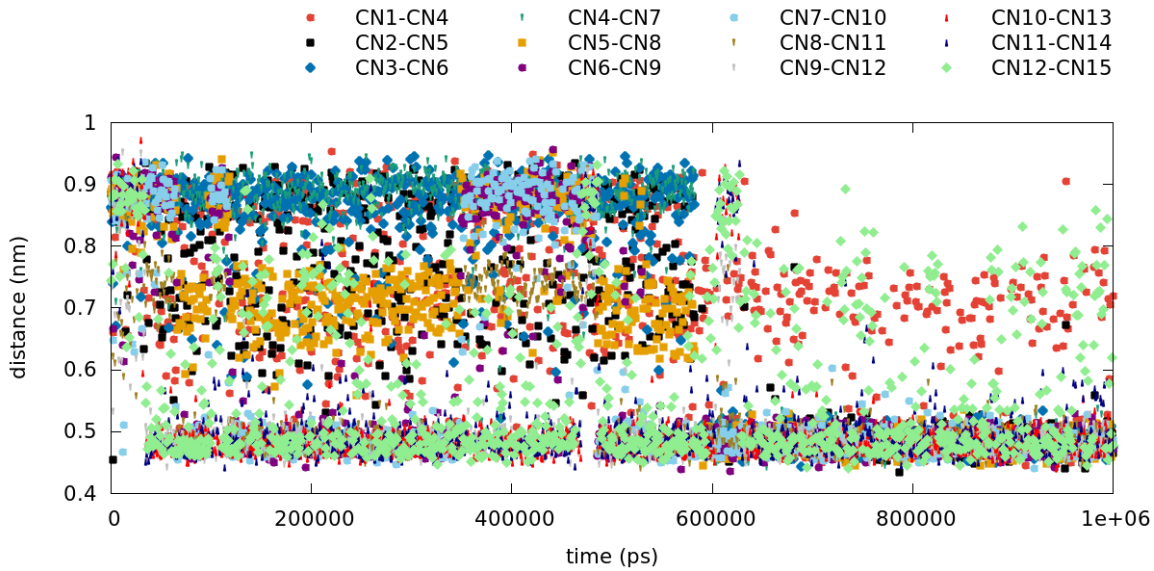

Figure 4.9: Hydrogen bonding distances (for $CN_i - CN_{i+3}$ pairs) versus time, for a single LK model simulated at the CG interface for $1\mu s$. Starting from an extended conformation close to the interface, LK adsorbs to the interface and folds back to an ideal helix adopting correct partitioning of hydrophobic/hydrophilic side chains.

Up until now 15.6 kJ/mol is the best value for hydrogen bonding E_{hb}. With this final set of tabulated potentials for hydrogen bonding, random behaviour of single LK in implicit solvent, and folding back of an extended structure in the presence of CG interface have been observed. Another measure for the strength of hydrogen bonding can be the end to end distance distribution of the helix. In order to represent that we used potential of mean force for pulling two ends of helix, which will be explained in the next session.

4.3 PMF based approach

In this section we have pulled the two ends of peptide away so that all the hydrogen bonds are separated. The potential of mean force for this pulling has been calculated for CG model with the obtained hydrogen bonding potentials from structure based approaches (E_{hb} 15.6 kJ/mol) and different E_{hb} values. Umbrella sampling simulation of pulling the two ends of atomistic helix (NME and ACE) has been done by Cahit Dalgicdir and will be used as our reference. For CG model the first and the last C_α (CA0 and CA15) are pull and reference groups. Pull group has been pulled away from the reference group in three dimensions in implicit solvent with pulling rate of 0.01 nm/ps and pull spring constant of 1000 kJ/$(mol.nm)^2$. Simulation with the initial conformation obtained from this pulling trajectory has been done and PMF curve has been calculated using weighted histogram analysis.

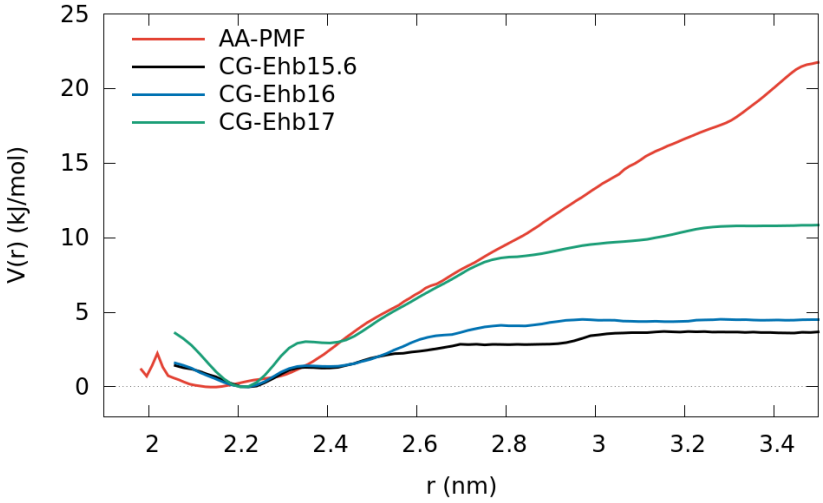

Figure 4.10: PMF curves for pulling the two ends of a helix in the CG model with different E_{hb} values in comparison with AA PMF reference. PMF curves has been shifted with the difference between pull and reference groups in AA and CG model (CA0-CA15 and NME-ACE distances). The slope between 2.2 and 2.8 nm is matched for E_{hb} 17 kJ/mol.

E_{hb} 17 has been chosen as the closest ϵ value for hydrogen bonding based on the PMF curves comparison in Figure 4.10. A simple calculation of the number of hydrogen bonds in each frame shows that there are approximately no hydrogen bonds

while pulling after 3.5 nm. (Figure 4.11) Thus the difference between AA and CG curves is associated with the weak backbone dihedrals used for CG model (*vacuum-excl.* dihedrals).

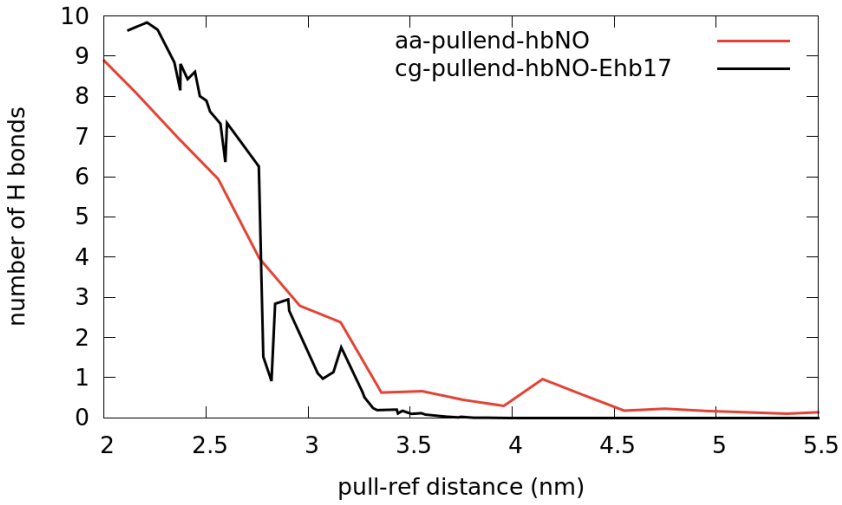

Figure 4.11: Number of hydrogen bonds versus end to end distance in umbrella pulling simulation of two ends of a helix. Curves has been shifted with the difference between pull and reference group in AA and CG model (CA0-CA15 and NME-ACE distances). After 3.5 nm there are approximately no hydrogen bonds in both cases.

Later in Chapter 6 we will see that electrostatics is playing an important role in CG model. By introducing the hydrophobic attraction for aggregates we will see its effect in the long range behaviour of PMF curves in CG simulations comparing to atomistic ones. In order to mimic the long range behaviour we need to take the electrostatics into account. Then it comes to the point that Lysine-Lysine repulsion also matters for the stability of helix. With 1.0 nm cutoff for coulombic interactions we hardly take this repulsion into account but with particle mesh Ewald calculation we have the negative effect of electrostatic repulsion on the stability of helix.

Umbrella sampling simulation of pulling the two ends of peptide has been done using PME electrostatics with Fourier spacing of 0.24. The resulting PMF curves are plotted in Figure 4.12. The slope in CG PMF curves decreased due to the repulsion between lysines but still with E_{hb} 17 kJ/mol, hydrogen bonding mimics the behaviour in AA reference. Difference in the curves slopes after 3 nm is due to using flexible

dihedral potentials obtained from the extended structure in *vacuum-excl.*

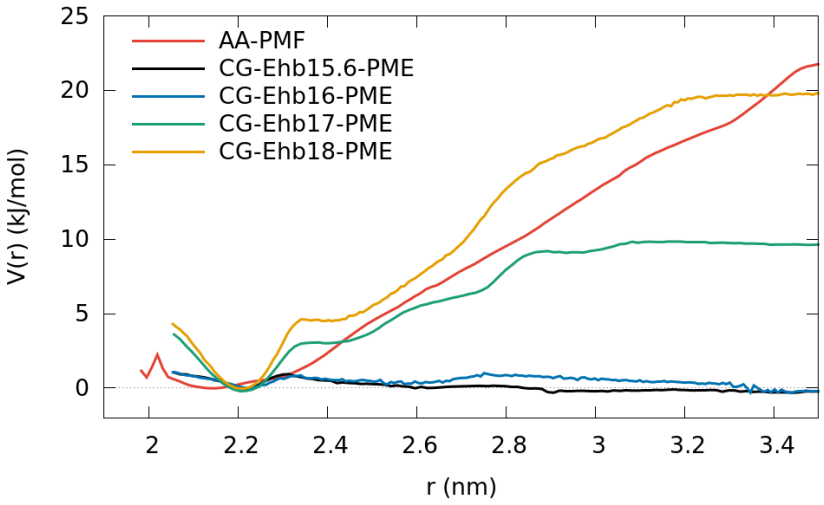

Figure 4.12: PMF curves for pulling the two ends of a helix in the CG model with different E_{hb} values in comparison with atomistic reference PMF in the presence of PME electrostatics. PMF curves has been shifted with the difference between pull and reference group in AA and CG model (CA_0-CA_{15} and NME-ACE distances).

From now on by switching to PME calculation for electrostatics we will show that this new hydrogen bond strength (E_{hb} 17 kJ/mol) also satisfies structure based approaches. We can see the frequent folding-unfolding events in a single LK simulation in implicit solvent. LK can adopt different conformations starting from the helix or extended structure. Ability of LK to fold back and make hydrogen bonds also support the structural based criteria.

By starting from the helix and extended conformations, single LK has been simulated in implicit solvent while PME electrostatics being used with Fourier spacing 0.24. Resulting time-line data of hydrogen bonding distances is plotted in the Figure 4.13 shows the random nature of helix in bulk water. LK is able to fold and unfold frequently during the course of the simulation.

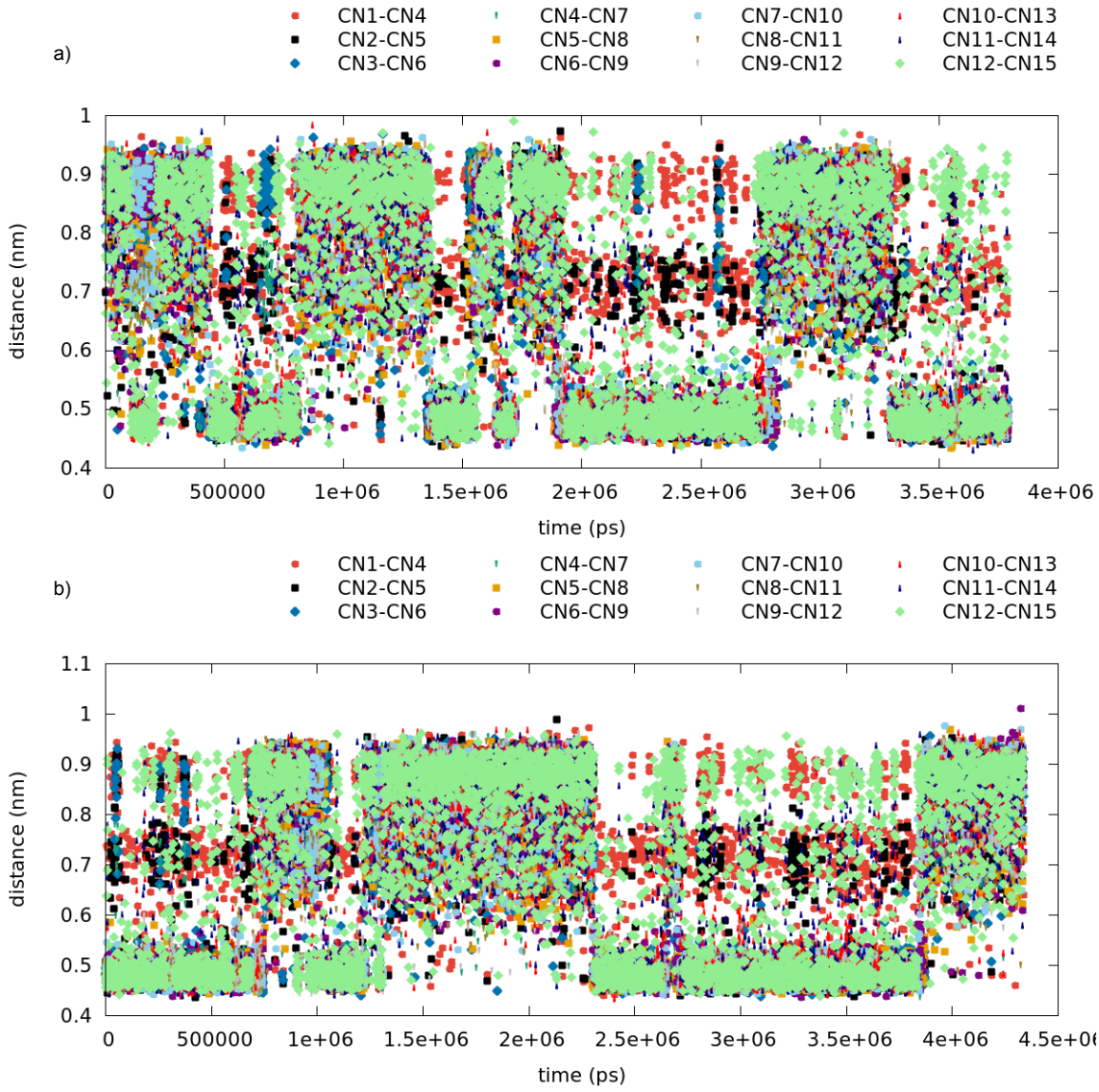

Figure 4.13: Hydrogen bonding distances ($CN_i - CN_{i+3}$) versus time, for a single LK model simulated in implicit solvent for more than $3\mu s$. E_{hb} has been set to 17 kJ/mol and PME electrostatic calculations have been applied to the CG model. Starting from an ideal α-helix (a) and an extended (b) conformations, frequent folding unfolding behaviour observed that verifies the ability of LK CG model to mimic the atomistic REMD behaviour.

4.4 Hamiltonian Replica Exchange

In addition to PMF and structural based approaches mentioned in previous sessions, Hamiltonian replica exchange has been used to enable the LK sample different conformations in a more efficient way. In HREX different replicas has been set based on different Hamiltonian. HREX simulation with 16 different replicas according to different, hydrogen bonding E_{hb} values has been run for $2\mu s$. In this method each replica has a different hydrogen bonding potential (E_{hb} values are different) and the idea is to excite the system with the help of strong hydrogen bonding potentials and at the same time let the system explore extremely low interaction potentials. As a result, high energy conformations including the ones that are kinetically trapped are replaced by the conformations with lower energies. 16 E_{hb} values between 14 and 17.75 has been used in a $2\mu s$ simulation and overall exchange probabilities have been calculated to be 84 percent. Resulting distributions for backbone dihedrals for some the replicas has been plotted in Figure 4.14.

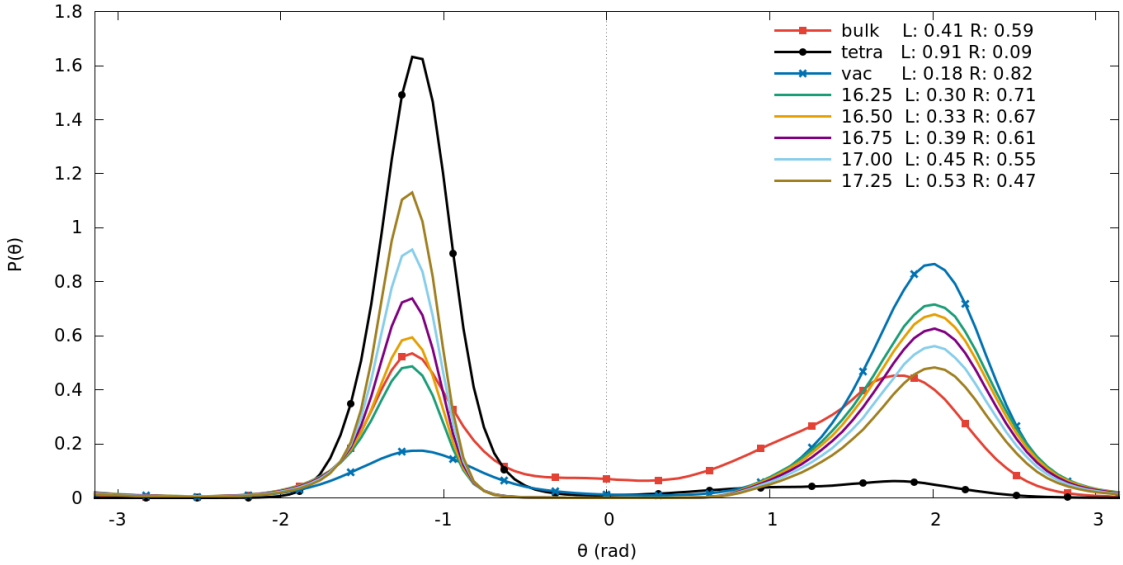

Figure 4.14: Backbone dihedral distributions for different replicas with different E_{hb} values ranging from 16.00 kJ/mol to 17.00 kJ/mol. L and R, are corresponding probabilities for left and right part of distribution.

Comparing the probability distributions with atomistic REMD, E_{hb} between 16.75 and 17 kJ/mol has been suggested which is close to the PMF based result (E_{hb} 17). In HREX calculation hydrophobic interactions are also defined and presented in our model. We will define and tune these interactions in the next chapters.

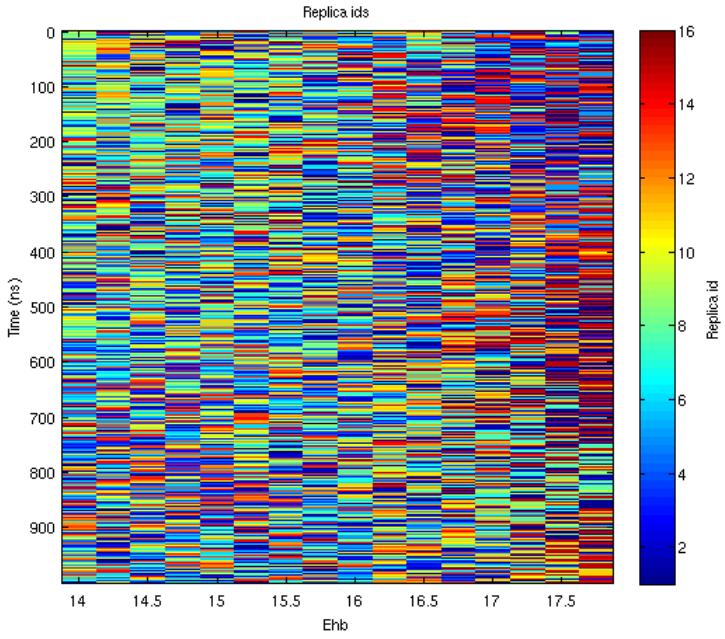

Figure 4.15: Matrix showing the exchanges between replicas with different E_{hb} for the HREX simulation in $1\mu s$. The color bar represents the initial replica IDs. The amount of mixing in HREX is correlated with the amount of mixing in this figure.

HREX removed the bimodal behaviour of probability distributions observed in backbone dihedrals (Figure 4.6) and smoothed the trend of these probability distributions. This method has enhanced the sampling of LK conformations by removing the energy barrier observed in dihedral probability distributions and even though we have not used beta hairpin conformation for our parameterization, we were able to sample this conformation. Due to leucine attractions half helix conformations have been also sampled using Hamiltonian replica exchange method.

Except HREX results, due to the lack of any hydrophobic interaction we were not able to observe any conformations resulting from hydrophobicity in this session. Half helix conformation where half of LK is in helical conformation and the rest is bended over the hydrophobic region to make a hydrophobic core, is not observed in implicit solvent. Behaviour of this model is also unknown in the presence of interface or other LK molecules. Now that different approaches suggests E_{hb} 17 kJ/mol for our model, we will define and tune other nonbonded interactions and study the behaviour of LK in the presence of all bonded and non-bonded interactions.

Chapter 5

NON-BONDED INTERACTIONS: HYDROPHOBIC INTERACTIONS AT THE MACROSCOPIC INTERFACE

In order to obtain the correct conformational behaviour of CG model at interfaces or in aggregates, one has to take the hydrophobicity of Leucines into account. We reproduced interface behaviour using the hard wall definition of Gromacs, and for aggregates we have used an attractive Lennard-Jones potential for Leucines to represent the hydrophobic forces.

We divided the hydrophobicity concept into two categories: Hydrophobicity in the presence of the interface, and hydrophobicity in aggregates. This division has been done since introducing simple attractive potential between leucines could be problematic in the presence of wall potentials. When peptide is at the interface and leucines are exposed to air or vacuum, they should not feel any extra attractions between themselves. We will find the parameters for these two interactions independently, and show the compatibility of parameters from thermodynamic based approaches with structural based approaches.

5.1 Structural based approach

Simple definition of a hard wall by Gromacs 4.5.6 which adsorbs leucines with LJ 6-12 potential and repels all the other beads with a generic repulsion failed to keep the helix stable at the interface. As expected, starting from the extended conformation, LK does not fold back and make hydrogen bonds in the presence of wall. Due to this

inability of wall to represent air-water interface, we have used a more complicated set of potentials to represent the soft interface. We need leucines to be attracted to the interface and when located at the interface, represent the correct partitioning of all residues. In other words, we need our charged lysines to be away from the interface because of their hydrophilic nature and backbone bead wondering around the interface location. To be consistent with the atomistic interface simulation, density profiles of three beads has been measured in both AA and CG simulations.

For wall-leucine attractive potential we used a modified LJ cosine potential (Eq 5.1) that is constant for the values before r_m (region after CG interface) and is zero for the values after cutoff (in water slab).

$$\begin{cases} U = -\epsilon & r \leq r_m \\ U = -\epsilon \left[\cos\left(\frac{\pi(r-r_m)}{2(r_c-r_m)}\right)\right]^2 & r_m < r < r_c \\ U = 0 & r_c \leq r \end{cases} \tag{5.1}$$

Considering r_m as the interface location, leucines get attracted to the interface when they are close enough and after passing through the interface and getting exposed to air, they would be in the favorable region. The question is how close the leucines can get to feel the wall attraction. One can set this distance by using different approaches. One way is to find the thickness of atomistic interface from density profile of water slab. (Figure 5.1)

Left side of this density profile is where LK has been located, that is why the slope is smoother. On the right side the thickness of the interface is measured to be 1.0 nm. If we take the average of min and max density, that would correspond to where the interface has been located (x = 6.77 nm and grey intersection). Then we can use the distance between the maximum density (x = 6.17 nm) and the interface location as our interface thickness (d = 0.6 nm).

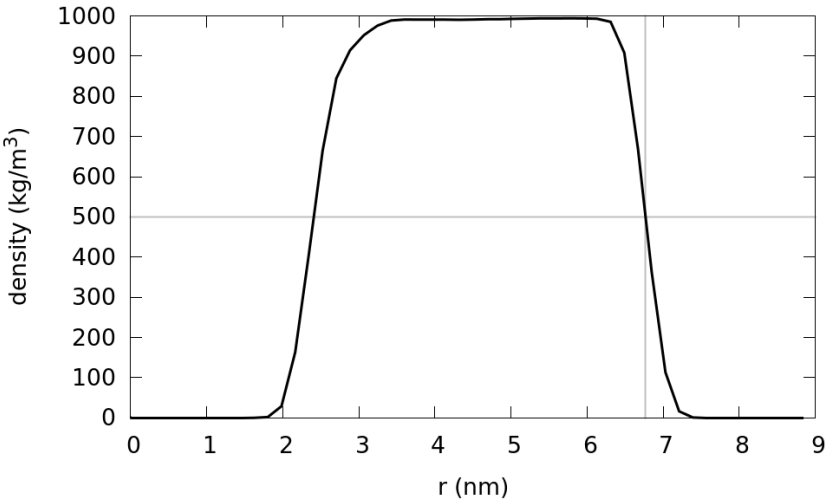

Figure 5.1: Density profile of water slab in interface AA simulation. The smoother slope on the left side of this curve is where LKα14 has been located. The grey intersection corresponds to the interface location. (between maximum and minimum density)

In our model we set our interface at 1 nm from the box end (hard wall of Gromacs) to be able to reduce the number of tabulated potentials used for different bonded and non-bonded interactions by setting multiple interactions to one table.

Having interface at 1 nm from the box end, attractive gain of potential has to start from 1.6 nm from the box end (1+d). We have initially used ϵ -10 for the wall-Leucine attraction and Figure 5.2 shows this attractive potential. Later we will discuss about how the determination of ϵ took place.

In addition to wall - leucine attractive potential we have added two generic repulsive potentials (Eq 4.2) to our model. One to keep lysines away from interface (lysine - wall repulsion) and the other one to keep all other beads from penetrating the Gromacs wall at the end of the simulation box. To mimic the correct distribution of beads at the interface, we compared the atomistic density profiles of Leucines, Lysines, and C_α. Helical LK with E_{hb} 15.6 has been simulated in implicit solvent, and repulsive potentials for interface has been tuned over several iteration to mimic the atomistic density profiles at the interface while E_w has been arbitrarily set to 10 kJ/mol. A

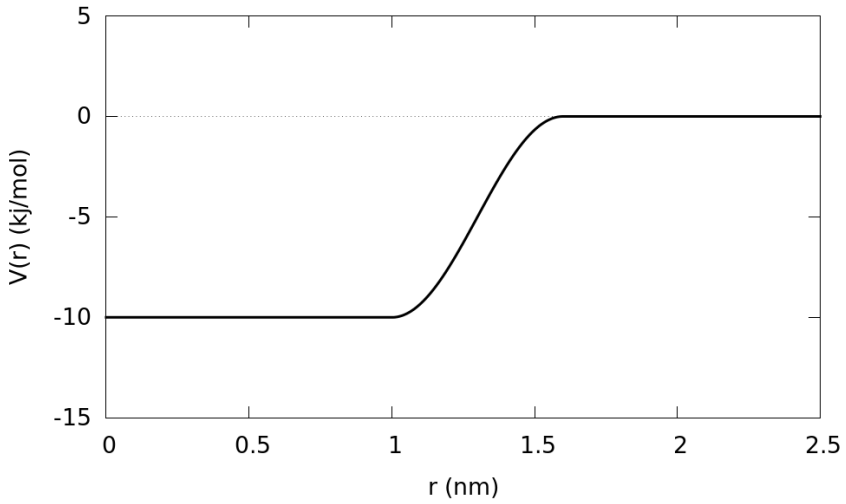

Figure 5.2: Interface - Leucine attractive interaction potential. Interface has been located at 1 nm from the box end. Leucines feel the hydrophobic interaction at 1.6 nm

comparison of these density profiles has been provided in Figure 5.3

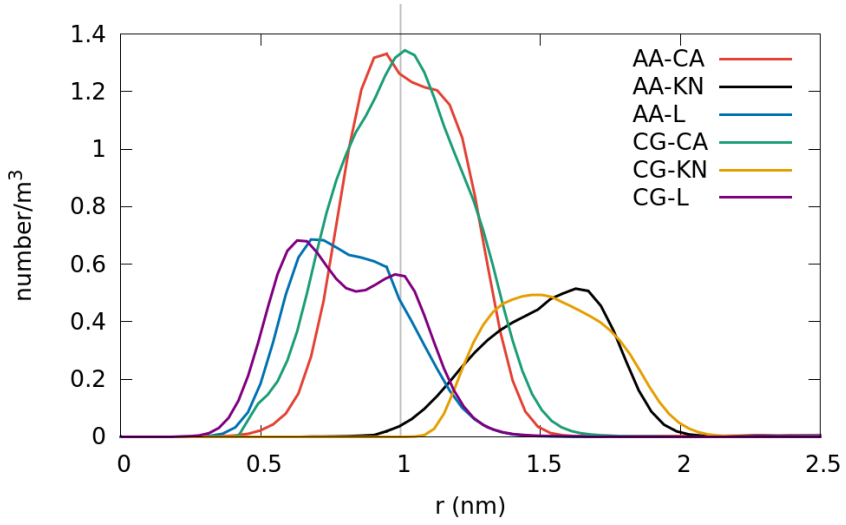

Figure 5.3: Density profiles of backbone, hydrophobic leucines and hydrophilic lysines at the CG interface in comparison with atomistic density profiles. E_w has been set to 10 kJ/mol, and E_{hb} is 15.6 kJ/mol

Schematic representation of interface and its involved interactions, has been illustrated in Figure 5.4. Based on the comparison between partitioning of hydrophobic, hydrophilic and backbone beads, following values have been obtained for tabulated

potentials involved at the interface:

- KN-wall repulsive potential: ϵ= 1.0 kJ/mol r_m=1.4 nm
- Others-wall repulsive potential: ϵ=10 kJ/mol r_m=0.495 nm

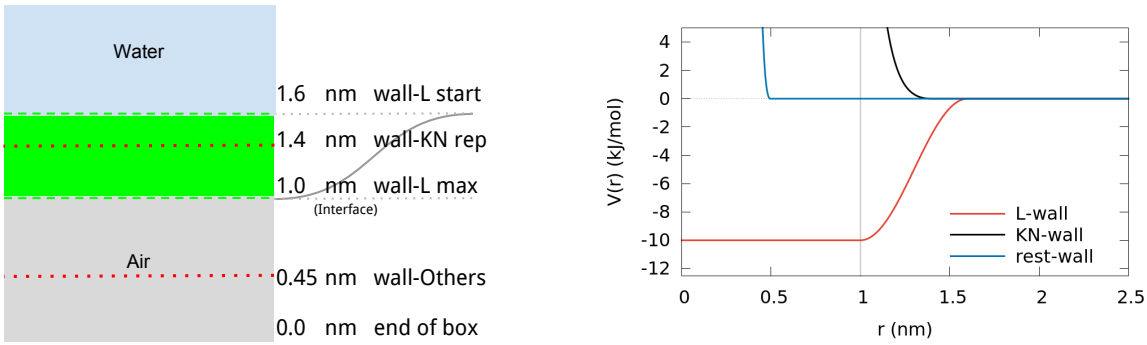

Figure 5.4: Interface interaction potentials for CG model (right) and their schematic representation (left). E_w has been arbitrarily set to 10 kJ/mol.

Using structural based approaches, we could not get an estimate for wall - Leucine ϵ value (E_w). It has been observed that with different values of 5,10 and 20 kJ/mol for E_w, generated density profiles are similar and AA atomistic behaviour for partitioning at the interface is still observable.

Even though with E_w 20 and 5 we have correct partitioning at the interface, our model failed to fold back to an α-helical structure while E_{hb} is 15.6 kJ/mol. LK stays extended for a 1μs simulation. Using E_w 10 kJ/mol, folding behaviour observed with E_{hb} 15.6 as shown in time data of $CN_i - CN_{i+3}$ distance values during the course of 1 μs simulation in Figure 4.9. Starting with an extended conformation it adsorbs to the interface and after 350 ns, then partially folds back to an α-helical structure up to 600 ns and eventually to an ideal helix.

5.2 Thermodynamic based approach

Umbrella sampling has been used to optimize the E_w value for wall - leucine interaction. Umbrella sampling simulation of pulling helix from interface has been done for all atomistic LK by Cahit Dalgicdir. We have started the CG umbrella sampling simulation with fixed hydrogen bonds for simplicity. A simple harmonic potential has been used

for hydrogen bonds so that LK stays helical and independent of hydrogen bonding potential. Our aim is to determine wall - L ϵ value (E_w) by matching PMF curves of CG and AA simulations.

A comparison of pulling PMF curves before taking PME calculation into account has been provided in Figure 5.5. The difference between using flexible hydrogen bond potential and fixed harmonic bonds while E_w is 10 kJ/mol suggests that peptide sacrifices some of its hydrogen bonds to lower the potential of mean force while pulling from the interface. With this ϵ value (E_{hb} 15.6) the cost of sacrificing hydrogen bonds are less than having interaction with wall. As a result we have to take this difference into account while we are using harmonic (fixed) bonds instead of a flexible LJ potential. Based on this difference and compared to atomistic PMF, E_w 20 kJ/mol is the proper value for wall potential.

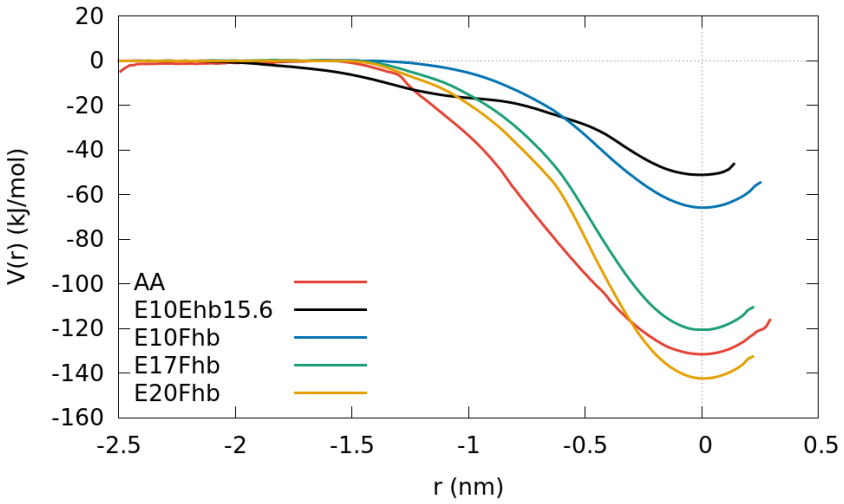

Figure 5.5: PMFs of pulling LK from the interface. CG model with different E_w has been compared to AA PMF. Fhb stands for fixed hydrogen bonds (using a harmonic potential for simplicity).

Using the new wall - L interaction potential (with E_w 20), free simulation of single LK at the interface has been run. Resulting density profiles and partitioning of hydrophobic, hydrophilic and backbone residues in helix at the interface are matched with AA interface results. Except the folding unfolding events for end residues which is also observable in AA case, helix is stable during the course of the simulation.

With the new wall - L strength (E_w=20 kJ/mol) and our old hydrogen bonding (E_{hb}=15.6 kJ/mol), LK failed to fold back to helix from an extended conformation at the CG interface in a 1μs simulation. Extended conformation attaches to the interface but hydrogen bonding is not strong enough to pass the corresponding energy barriers. This suggest that we need stronger hydrogen bonding potential, which can parse over the local regions for hydrogen bonding and passes through the high energy barriers due to repulsive potentials.

Up to now, PME electrostatic has not been employed in our model. Using PME, and hydrogen bonding representation with VS obtained from the previous section (E_{hb}=17 kJ/mol), the potential of mean forces for pulling peptide from interface are reproduced. Resulting PMF curves have been shown in Figure 5.6.

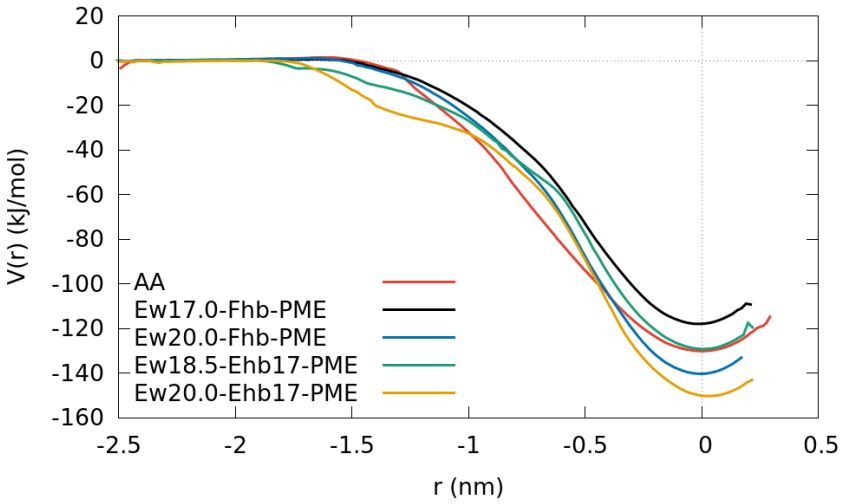

Figure 5.6: PMFs of pulling LK from the interface in the presence of PME electrostatic calculations. CG model with different E_w has been compared to AA PMF. Fhb stands for fixed hydrogen bonds. Among the different strength for wall - L attractive potential, E_w 18.5 result, perfectly matches with the AA PMF

Taking PME into account, and with E_{hb} 17, proper E_w value for wall potential is 18.5 kJ/mol. Here the difference between the fixed hydrogen bond and the flexible one for E_w 17 shows the cost of sacrificing hydrogen bond is more than interacting with wall. LK favors being helical more than touching interface in longer distances by breaking its hydrogen bonds. As expected with E_w 18.5, hydrophobic, hydrophobic

and backbone beads have correct partitioning at the interface as can be seen in the density profiles of corresponding beads for a 1 μs simulation of LK with E_{hb} 17 at the interface in Figure 5.7.

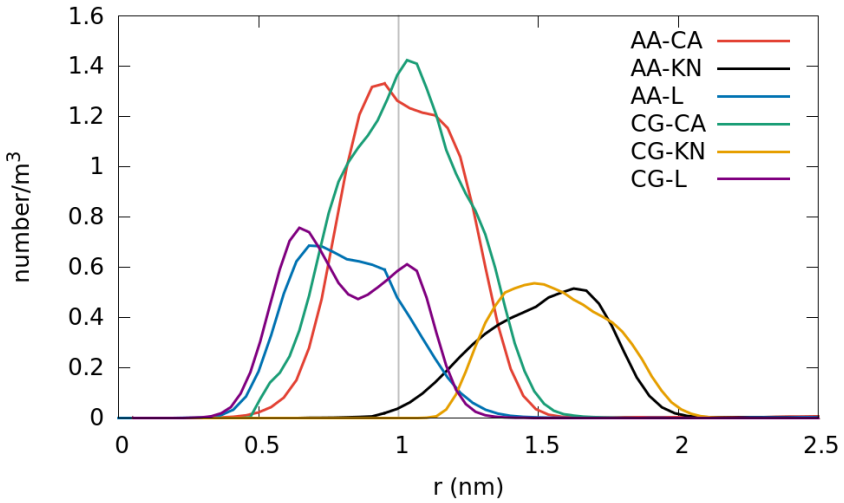

Figure 5.7: Density profiles of backbone, hydrophobic Leucines and hydrophilic Lysines at the CG interface in comparison with atomistic density profiles. PME electrostatics has been employed in our system. E_w has been set to 18.5 kJ/mol, and E_{hb} is 17 kJ/mol

Using E_w 18.5, E_{hb} 17 and PME for electrostatics, LK has been simulated at the interface starting from an extended conformation for 1 μs. Extended LK folds to helix at the interface similar to the folding behaviour observed in atomistic simulations. The time-line data of hydrogen bonding distances together with number of hydrogen bonds versus time shown in the Figure 5.8 display the folding process.

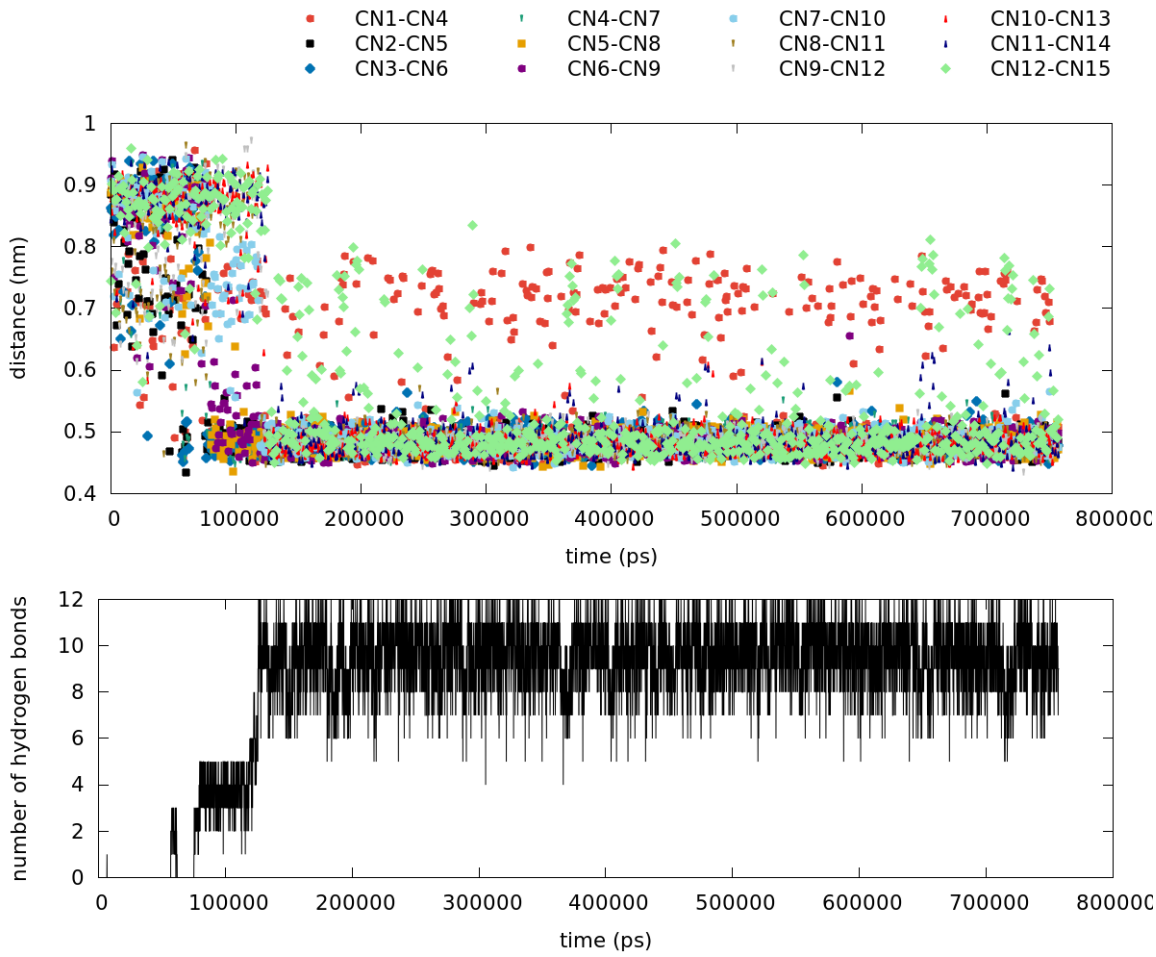

Figure 5.8: Hydrogen bonding distance $CN_i - CN_{i+3}$ versus time (top) for a single LK model, folded to an α-helix in the presence of CG interface in implicit solvent and while PME electrostatics has been taken into account. Number of hydrogen bonds versus time (bottom) also shows this transition to helix. E_w has been set to 18.5 kJ/mol, and E_{hb} is 17 kJ/mol.

Thermodynamic based (PMF) and structure based (partitioning) criteria have been satisfied with this set of potential for wall interactions which has the E_w of 18.5 kJ/mol. Hydrogen bonding potential is also in agreement with both approaches. In reality the attraction between leucines is due to the water, so at the interface leucines should not feel this hydrophobic attraction. The options that we have are either turn them off completely or leave them as they are. In the next section we will introduce this hydrophobic interaction and see the behaviour of extended conformation at the

interface with the presence of leucines and finally choose the second option and leave Leucine interaction presented in the CG model.

Current LK model is able to show the all atomistic REMD behaviour in bulk and to fold back to an α-helical structure in the presence of a CG interface. In other words, LK model with the current parameters is transferable to different conformation in bulk and at the interface. Up to now we solved the transferability issue for our CG model with these new set of potential. But it still fails to show any aggregation behaviour because of the lack of any potential representing hydrophobicity of leucines in bulk. In the next Chapter we will define this potential, and show the agreement of structural based approach with thermodynamic approaches for parameterization to reproduce the atomistic LK behaviour.

Chapter 6

NON-BONDED INTERACTIONS: HYDROPHOBIC INTERACTION AT THE MOLECULAR INTERFACES

We saw that hydrophobicity of Leucine side chain leads to partitioning at the interface, and hydrogen bonding induces folding at the interface. In the absence of any hydrophobic interactions between leucines, single LK failed to show conformations which contain a hydrophobic core, like half helix. In addition to that several LK molecules are moving freely in the box and no signs of aggregation has been observed as expected. In this session we are trying to define and tune hydrophobic interaction that leads to aggregation and emergence of a hydrophobic core. For this purpose radial distribution function of leucines has been measured in tetramer atomistic simulation.

In both experimental and theoretical studies, it is known that among different aggregate sizes, tetramer is the most stable one [1, 25]. Thus we use radial the distribution function of leucines in atomistic tetramer to parameterize leucine-leucine interaction potential. A modified Lennard-Jones with cosine term (Eq 4.1) fitted to the Boltzmann inversion of RDF (Figure 6.1) and r_m is set to 0.507 nm. Cutoff (r_c) for this LJ potential has been set to 0.7 nm so that it avoids the second peak in RDF.

In order to avoid the immediate collapse of the LK model we need to add exclusions for the first and second nearest L neighbors since their equilibrium distances in helical conformation are less than the cutoff. Now that the minimum for LJ cosine potential has been set to 0.507 nm, we will obtain the proper value for L-L E_L using thermodynamic based approaches, and check the compatibility of resulting parameters with structure based criteria.

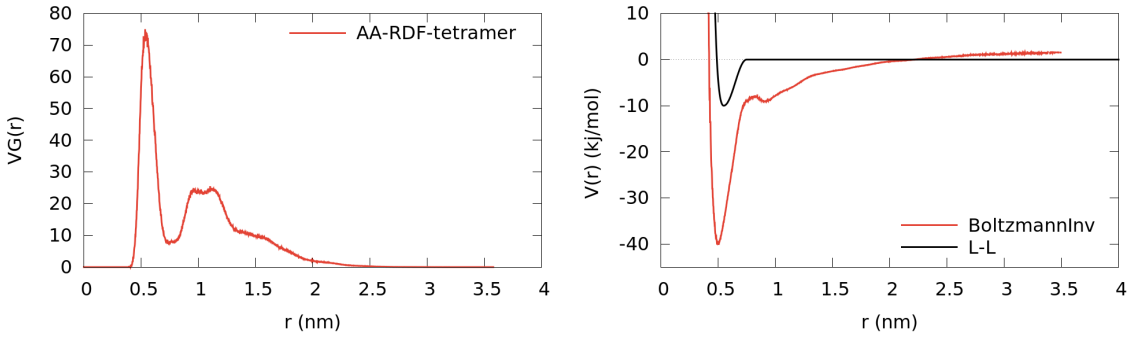

Figure 6.1: RDF of Leucines in atomistic tetramer simulation (left), and tabulated potential for L-L interactions on top of Boltzmann inversion of AA RDF (right). A modified LJ potential with cosine term has been used for L-L and it has an arbitrary ϵ value (E_L) of 10 kJ/mol in this figure.

Dimer pulling umbrella sampling simulation has been done to tune the E_L so that we can mimic the atomistic PMF curve. Similar to pulling from interface, hydrogen bonds are fixed with a harmonic potential for simplicity. Center of the mass of the backbone beads in one of the peptides is pulled away from the one the other LK with the pull rate of 0.005 nm/ps. Initially we did not have PME electrostatics defined in our model. Resulting potential of the mean forces are plotted for E_L 7 and E_L 8.5 kJ/mol in Figure 6.2.

Similar to what we have observed from pulling LK from the CG interface, presence of hydrogen bond increases the minimum value of PMF and make it smoother by sacrificing hydrogen bond upon increasing dimer-dimer distance. Due to this comparison 8.5 could be the proper value for E_L.

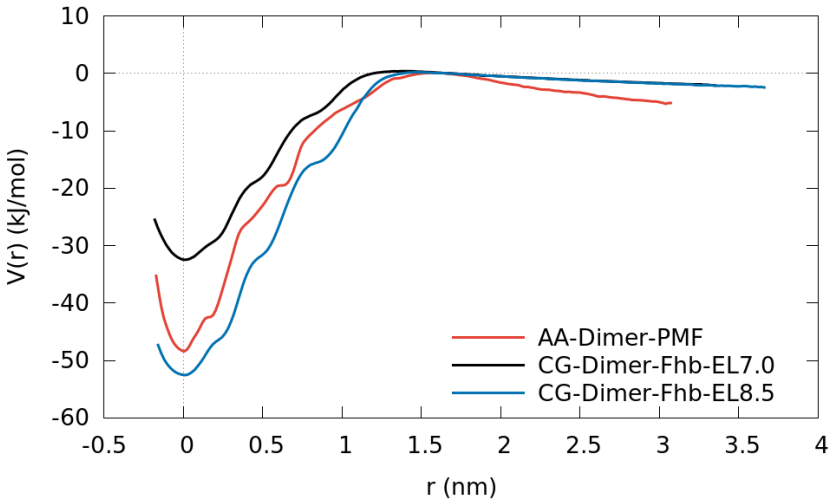

Figure 6.2: PMFs for separating one LK from dimer aggregate for different E_L values. The depth of PMF when E_L is 8.5 kJ/mol is close to AA reference PMF.

Since equilibrium distance of KNs in helix are around 1.0 nm, they barely feel electrostatic repulsion from other lysines while the short range coulombic calculations of electrostatics is presented in our model. As a result, lack of long range electrostatics effects long range behaviour (slope) of dimer pulling PMF curve as can be seen in Figure 6.2. The slope on the right side of AA PMF curve is due to the long range interactions which we are missing in our model. PMF curves reproduced in the presence of counterions also do not show changes regarding the long range behaviour of dimer. Thus to get a better resolution of parameters for hydrophobic interactions, we will re-tune the parameters after taking PME into account.

6.1 Particle Mesh Ewald electrostatics

Difference of slope in dimer pulling PMF curves (Figure 6.2), between CG and AA suggests that we need to take the long range electrostatics into account. Single LK has 6 charged lysines (KN), thus 6 CL counterions have been added to the system to maintain system neutrality.

Radial distribution function for KN and CLs has been measured for single peptide

in bulk atomistic simulation and compared to CG RDF of these pairs in Figure 6.3. Then we defined a repulsive interaction potential (Eq 4.2) between KNs and CLs to prevent Lysines and counterions getting too close to each other. Radial distribution functions for CL and KN has been plotted in Figure 6.3 before and after adding this potential. Epsilon for this repulsive potential is 10 kJ/mol and r_m has been set to 0.358 nm.

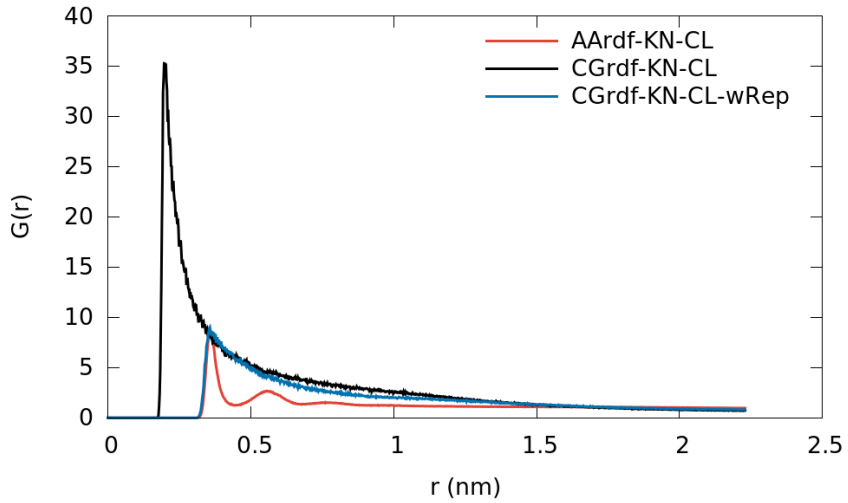

Figure 6.3: RDFs for KN and CL beads, for single LK in atomistic bulk water (red), and for CG model without introducing repulsion between KN and CL (black) and with the repulsive potential for these pairs (blue). In CG model we lose the structure observed in atomistic RDF due to the absence of CL ions' hydration layers.

The structure that we see in atomistic RDF (red curve in Figure 6.3) is due to the hydration layers of CL ions. Since water molecules are closely bonded to these ions we have the second and third peaks in AA RDF. But due to the absence of water molecules in our model we do not have such interactions and resulting RDF is structureless.

To increase the speed of simulation, actual time length of simulations with single LK using particle mesh Ewald with Fourier spacing of 0.12, 0.24, 0.36 has been compared. Optimal Fourier spacing of 0.24 while using two threads will be used for the PME simulations in LK coarse-grained model.

Using PME enabled us to mimic the correct behaviour of dimer in long ranges.

The slope of PMF curves in long range are in agreement with the atomistic one. Taking PME into account resulting PMF curves became smoother and the minima has been shifted up due to lysine - lysine repulsions. In Figure 6.4 we can clearly see the excellent agreement between E_L=8.5 kJ/mol and the atomistic PMFs.

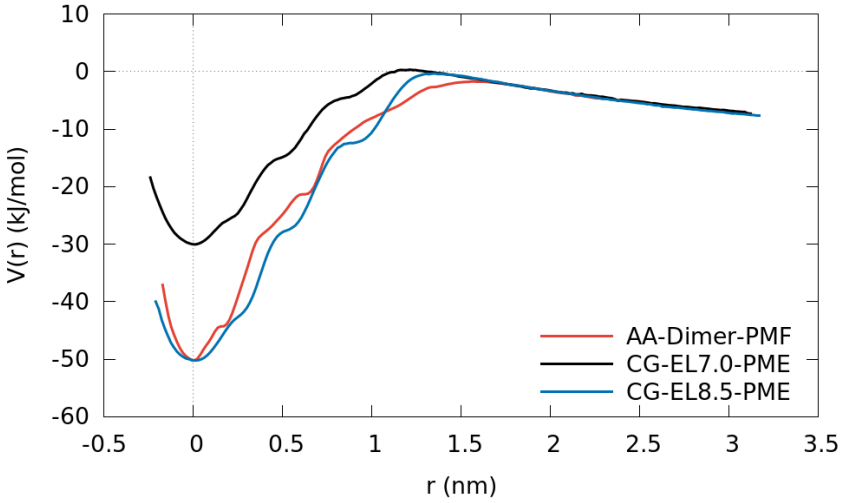

Figure 6.4: PMFs for separating one LK from dimer aggregate for different E_L values in the presence of PME electrostatics. The slope in long range of PMF were mimicked by using PME. The depth and form of PMF when E_L is 8.5 kJ/mol is in agreement with AA reference PMF.

All the dimer pulling simulations have been done using an anti-parallel dimer configuration. Figure 6.5 shows the comparison of dimer pulling PMFs for parallel and antiparallel orientations. As expected the difference between these curves is negligible and they both mimic the atomistic PMF curve.

Now that E_L 8.5 is found to be the closest value for ϵ of L-L attractive potential, instead of harmonic bonds, we can use finalized set of parameters for hydrogen bonding and reproduce the potential of mean force for the dimer dimer pulling. E_{hb} has been used for both LK peptides in dimer pulling simulation and resulting PMF curves have been plotted in Figure 6.6. In spite of the difference around the equilibrium distance, LK is able to produce a close curve to AA PMF curve.

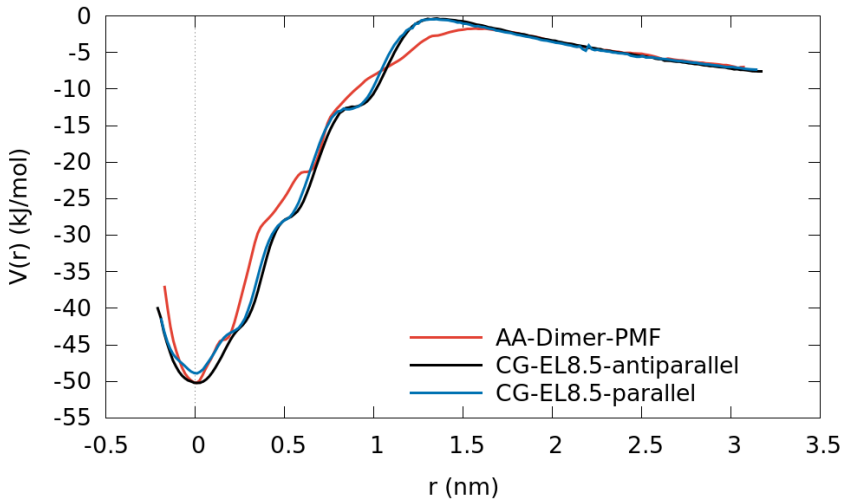

Figure 6.5: PMFs for separating one LK from parallel and antiparallel dimer orientation for E_L 8.5 kJ/mol and in the presence of PME electrostatics. The depth of PMFs are the same for both orientations which suggests that they are equally preferable.

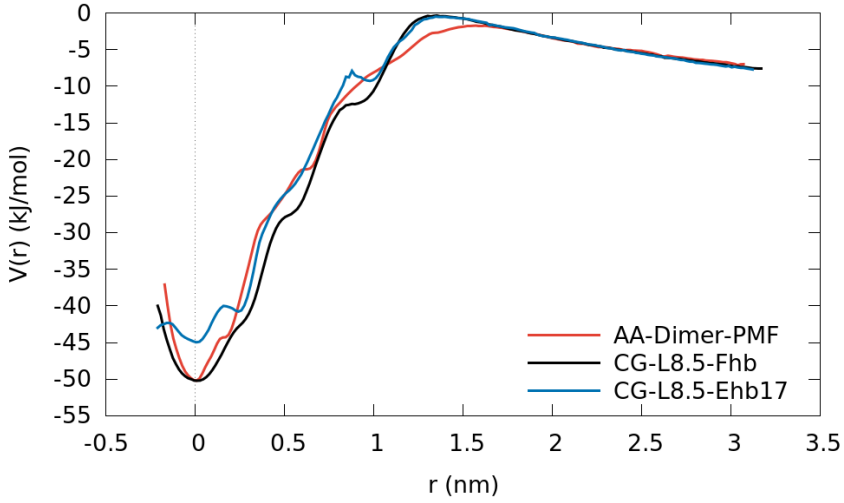

Figure 6.6: PMFs for separating LK from dimer for E_L 8.5 kJ/mol and in the presence of PME electrostatics. Hydrogen bonding interactions with E_{hb} 17 kJ/mol is used instead of harmonic bonds. The difference around equilibrium is due to the shift in COM of backbone when end residues are extended.

The difference around minima of PMF is due to the shift in the center of mass of the backbone in closer distances as shown in the Figure 6.6. One of the peptides is partially extended from the end residues and wrap the other LK to make a hydrophobic core. Figure 4.7 displays and snapshot of this dimer configuration.

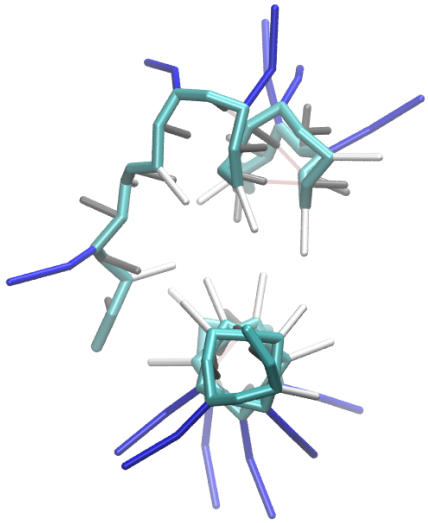

Figure 6.7: Snapshot taken from dimer pulling simulation while the dimer-dimer distance is around the equilibrium (minima in PMF curve). One of the LKs (top) is extended partially and alters COM of backbone beads. White beads are hydrophobic leucines and the blue ones are lysines, backbone is colored with cyan.

A distance-angle correlation plot has been obtained from the free simulation of dimer in implicit solvent for 500 ns is shown in Figure 6.8. Then compared to atomistic correlation plot as our structural based criterion to confirm the agreement with atomistic results. E_{hb} 17 has been used for hydrogen bonding and E_L 8.5 for leucine hydrophobic interaction. The distances in the correlation plot are between the center of mass of the backbone beads of peptides without taking into the consideration the first and last three residues. For the angle calculations, again the first and last three residues have been discarded and a vector was drawn between the remaining backbone beads of the second three residues from the beginning and second three residues from the end. the angle reported is the angle between these vectors.

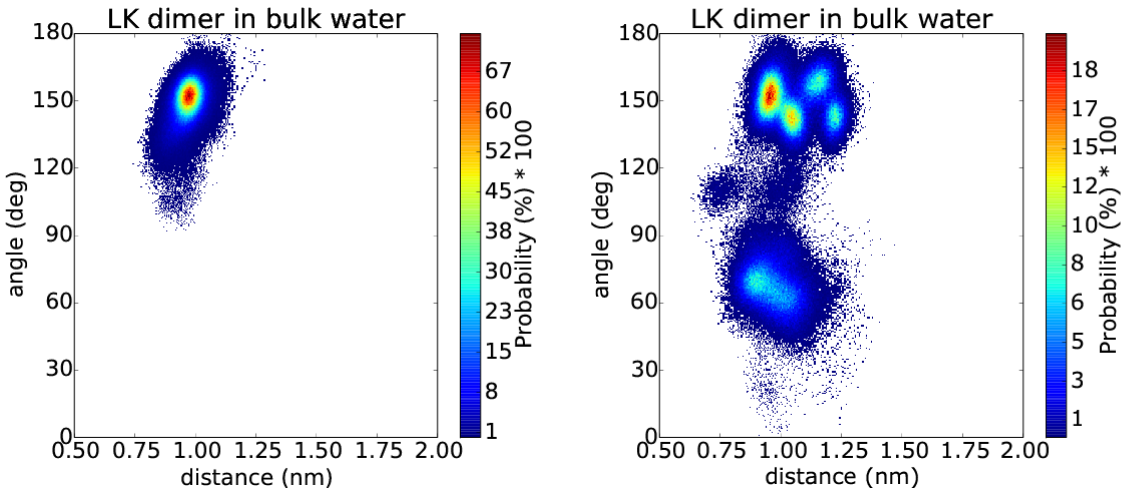

Figure 6.8: Distance versus angle correlation plots for LK dimer in atomistic (left) and CG (right) simulations. E_L 8.5 has been used together with E_{hb} 17 for CG dimer in implicit solvent. The bright spot corresponding to antiparallel configuration is in excellent agreement with atomistic results.

There are two equilibrium states in atomistic correlation plot which depend on starting conformation. Starting with antiparallel conformation the dominant orientation of the molecules in dimer would be parallel and starting with parallel we will have the reverse orientation for the dominant conformation. If the starting conformation is antiparallel, the dominant orientation of the molecules in the dimer would be parallel and vice versa . We have started with antiparallel orientation and bright spot corresponding to the maximum probability in coarse-grained correlation plot showed that it is in excellent agreement with atomistic dimer. Similar to atomistic dimer, in the CG model going back and forth between two parallel and antiparallel regions has been observed.

Other bright spots in CG correlation plot is due to different arrangements of LK to benefit from the COM and Leucine attractions at the same time. Distance versus time has been plotted with the corresponding snapshots showing these arrangement in Figure 6.9. The differences in distance is due to two different orientations of dimer: One is shifting up and down along the dimer axis and the other is when hydrogen

bonds in end residues are broken and bend over the hydrophobic core which changes the COM slightly.

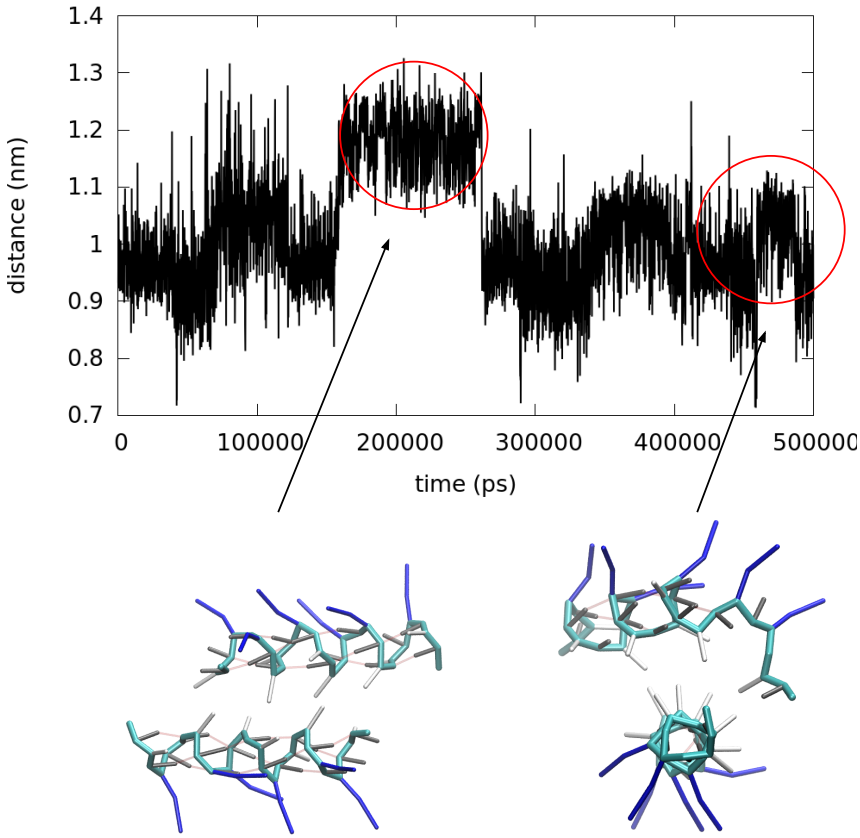

Figure 6.9: Distance versus time between the center of mass of the backbone beads of peptides without taking into the consideration the first and last three residues. E_L 8.5 has been used together with E_{hb} 17 for CG dimer in implicit solvent during a 500 ns simulations. Snapshots for different configurations adopted by LK dimer have been attached to their corresponding places. These configurations bring out the other bright spots observed in the CG correlation plot in Figure 6.8

The other structural quantity we have used to confirm the agreement of hydrophobic interactions in CG model with atomistic behaviour is radial distribution function of leucines. We have constructed the leucine-leucine interaction potential using the location of the first peak in leucine RDF from atomistic tetramer (Figure 6.1). Even though the attractive Lennard-Jones potential with cosine term has been constructed with the cutoff 0.7 nm to avoid the second peak of RDF, resulting RDF in Figure 6.10 shows that our model is able to sample atomistic RDF correctly.

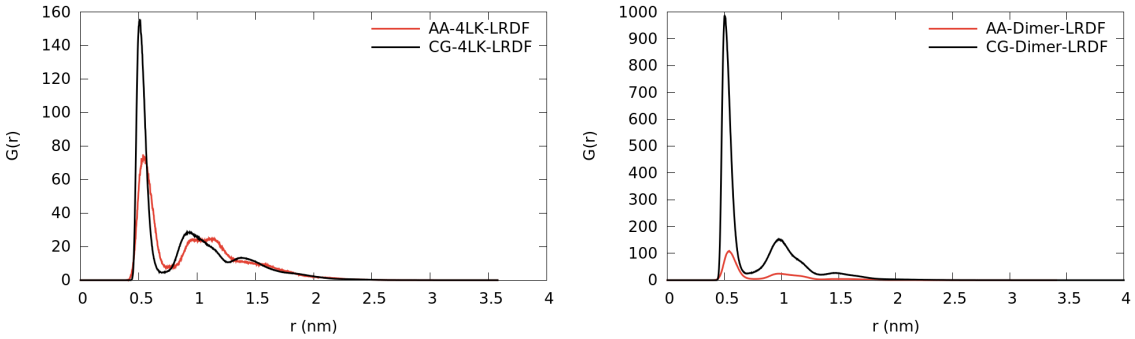

Figure 6.10: RDF comparison for tetramer aggregate in AA and CG model (left) and dimer RDF comparison (right). Even though we have not used the second peak of AA RDF to parameterize hydrophobic interaction, CG model has sampled it correctly.

The height difference between the major peaks in CG and AA RDFs suggests that the coarse-grained one with E_L 8.5 kJ/mol has higher attraction in comparison to AA. If one tunes the leucine attraction so that CG RDF mimics the height in atomistic RDF, the difference in PMF curves of AA and CG cases will be dramatically different (Figure 6.4).

The first and second peak in resulting tetramer CG RDF is slightly shifted to smaller values compared to atomistic RDF. This difference is known to be a normalization issue. Leucine-leucine interactions in tetramer are nonspherical objects, thus one has to consider the effects of using different normalization factors of r^2 or 1 (Figure 6.11). Since the difference in minima location of Boltzmann-inverted RDF curves is negligible we have used r^2 normalization factor which is generally for spherical distance distribution calculations.

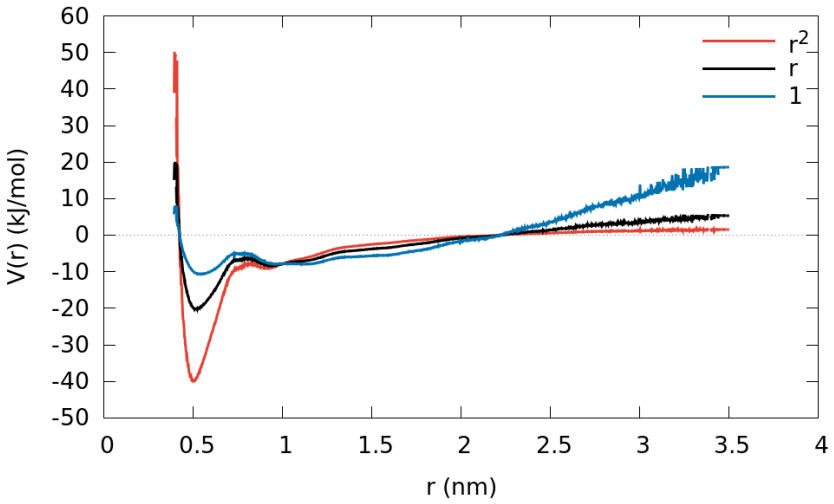

Figure 6.11: Comparison between Boltzmann inversion of RDF using different normalization factors. Different factors result in different minima locations.

Even though resulting parameters for leucine attraction obtained from potential of the mean force is in agreement with the correlation plot, due to decrease in DOF in coarse-graining we cannot capture all the features and conformations. Now that we have tuned all the bonded and non-bonded parameters, in the results section LK molecule will be tested with the presence of all these interactions to confirm the transferability of LK model and its agreement with atomistic behaviour.

Chapter 7

CONCLUSION

7.1 Single LK in implicit solvent

In this study we proposed an implicit solvent coarse-grained model for LKα14 that can mimic the behaviour of the molecule in three different environment; in bulk at the interface and in aggregate systems. LK behaviour is in agreement with the atomistic molecules behaviour in these environments. Atomistic REMD results in bulk were mimicked using the current set of potential for hydrogen bonding and hydrophobic interactions. In implicit solvent LK model adopts a variety of structures (Figure 7.1) observed in atomistic REMD simulation, including α-helix, extended conformation, partial helix and β-hairpin. This behaviour of CG model can be supported by employing Hamiltonian replica exchange method for replicas with different hydrogen bonding potentials.

This CG model is able to show frequent folding unfolding events and passes through energy barriers corresponding to hydrogen bonding in implicit solvent. Single LK with all the non-bonded interactions presented has been simulated in implicit solvent starting from the helical and extended conformations to confirm the ability of LK to fold, unfold and adapt different conformations in implicit solvent. Figure 7.2 shows these conformational transitions starting from helix and extended conformations.

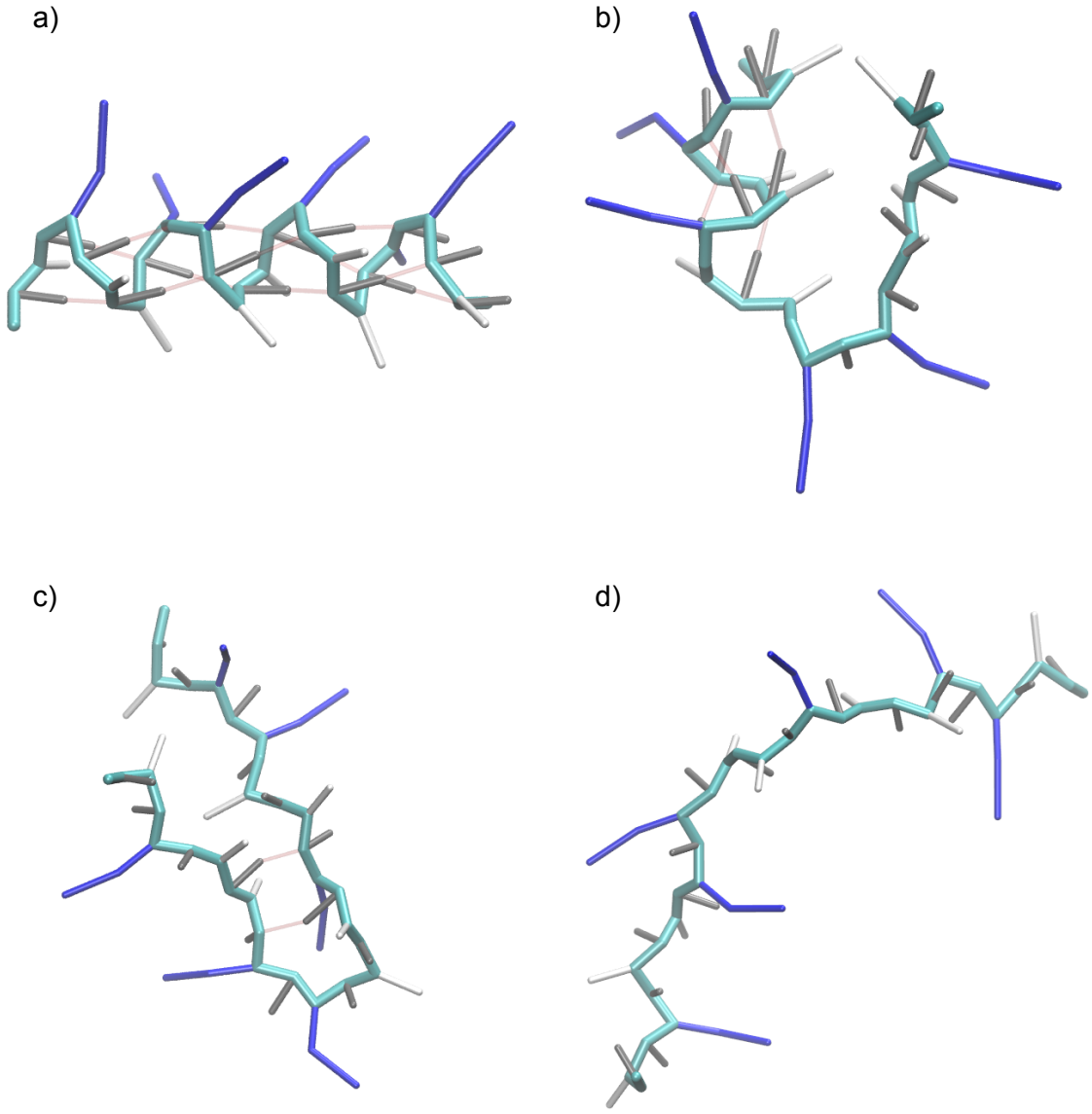

Figure 7.1: Different secondary structures sampled by HREX for LK model in implicit solvent: α-helix (a), half-helix (b), β-hairpin (c) and extended conformation (d).

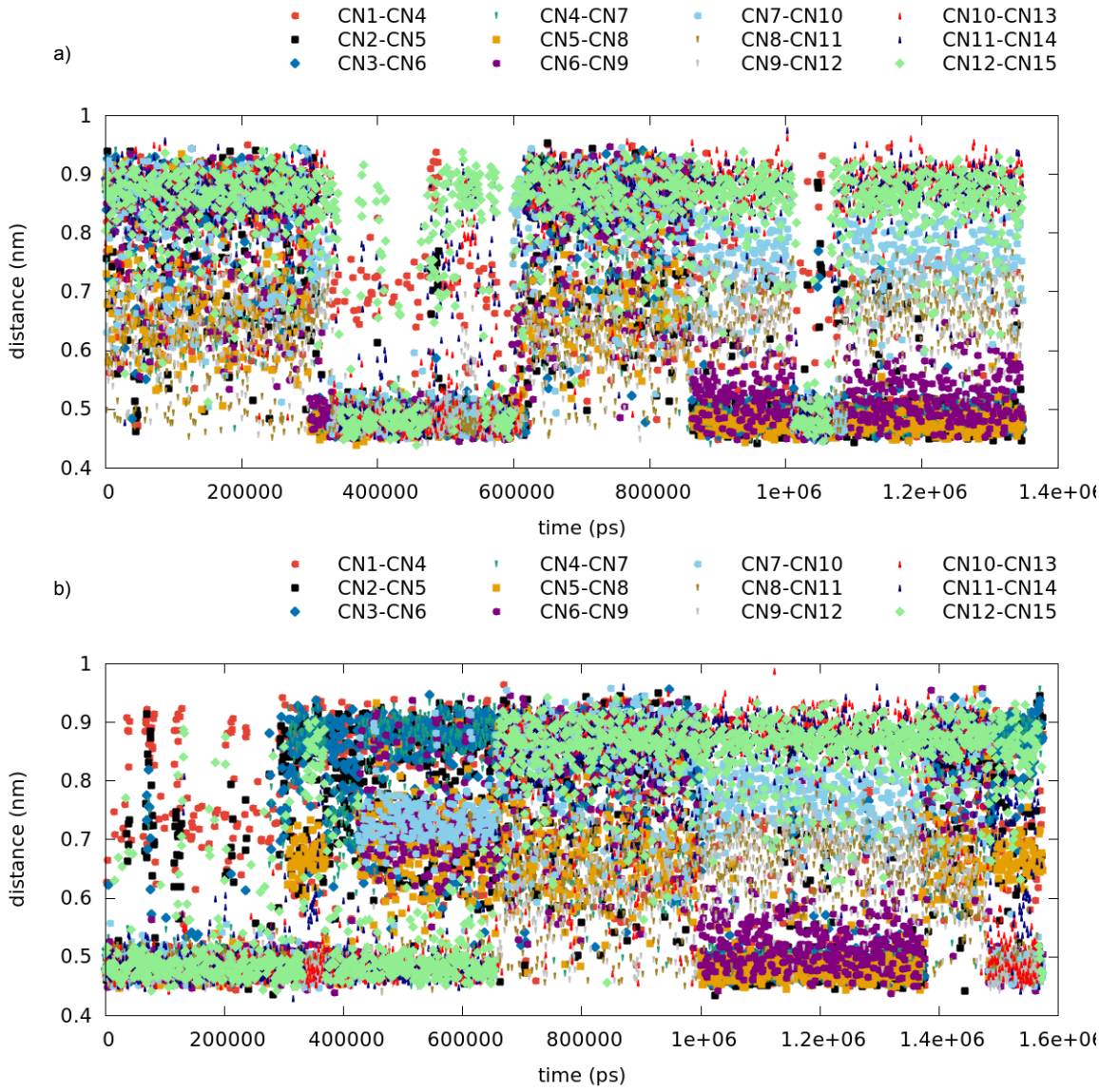

Figure 7.2: Hydrogen bonding distances ($CN_i - CN_{i+3}$) versus time, for a single LK model simulated in implicit solvent for more than $1\mu s$. E_{hb} has been set to 17 kJ/mol, E_L to 8.5 kJ/mol and PME electrostatic calculations have been applied to the CG model. Starting from an extended conformation (a) and an ideal α-helix (b), frequent folding unfolding behaviour observed that verifies the ability of LK CG model to mimic the atomistic REMD behaviour.

7.2 Single LK at the macroscopic interface

At the CG interface the disordered structure of LK folds into a stable α-helix and remains folded for the rest of simulation ($1\mu s$). The correct partitioning of hydrophobic/hydrophilic side chains has been induced by the CG interface defined in our model. Extended conformation has been simulated in the presence of all the non-bonded interactions, close to the CG interface for $1\mu s$ and behaviour of LK during the course of simulation has been reported in Figure 7.3.

7.3 LK at the molecular interfaces

The aggregation of LK model in implicit solvent has been investigated. Unlike the lack of a unique structure for LK in implicit solvent, in the presence of a molecular interface, extended conformation folds to an α-helix and adapts similar partitioning observed at the CG interface, i.e. hydrophobic leucines form a hydrophobic core and charged lysine side chains face the water molecules. Extended conformation has been simulated in bulk close to a helix for 500 ns and the hydrogen bonding distance $CN_i - CN_{i+3}$ versus time has been plotted in Figure 7.4.

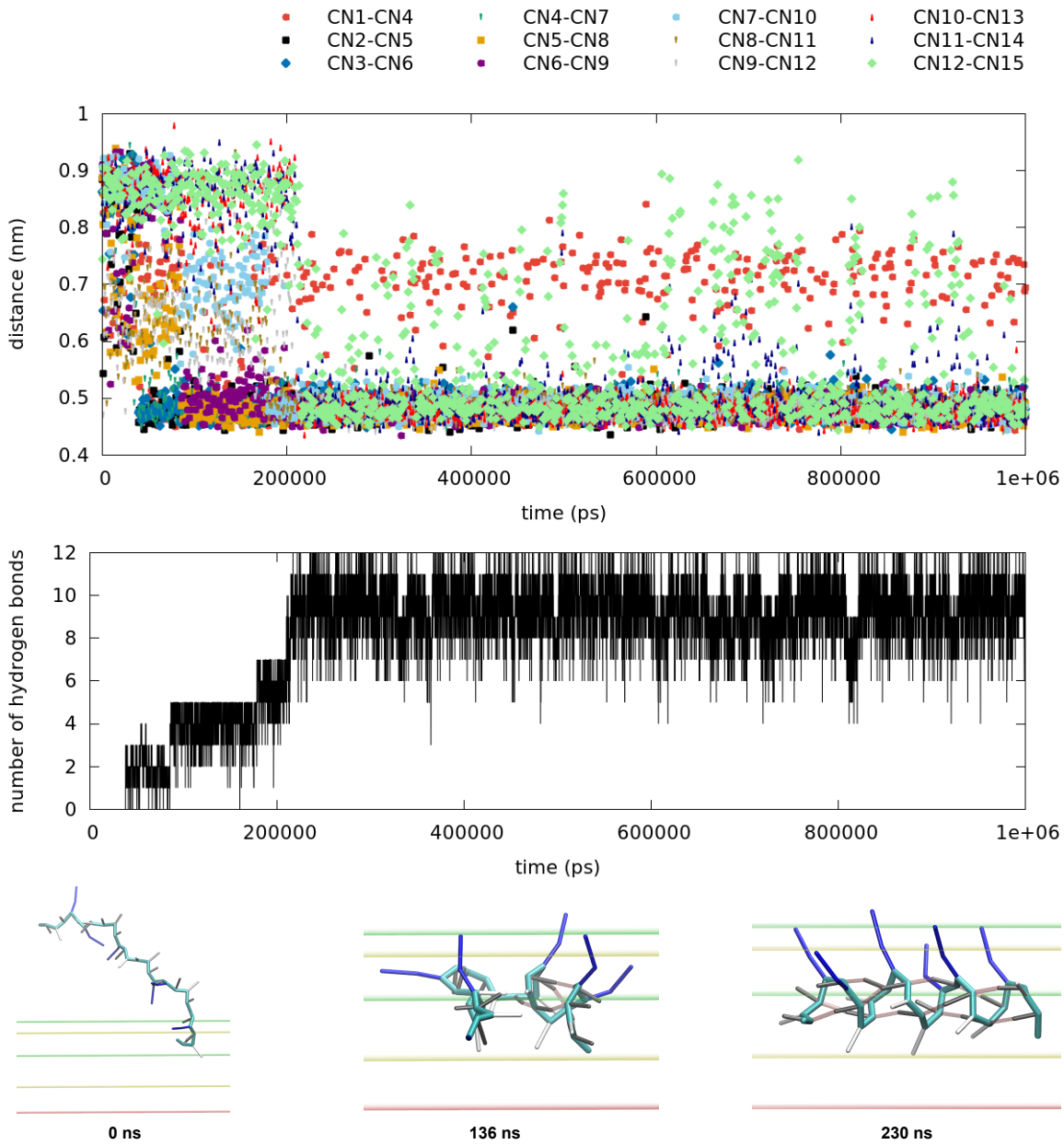

Figure 7.3: Hydrogen bonding distance $CN_i - CN_{i+3}$ versus time (top) for a single LK model, folded to an α-helix in the presence of CG interface in implicit solvent and while PME electrostatics has been taken into account. Number of hydrogen bonds versus time (bottom) also shows this transition to helix. E_w has been set to 18.5 kJ/mol, and E_{hb} is 17 kJ/mol and E_L 8.5 kJ/mol.

It is known from the experimental and theoretical studies, that tetramer is the most stable aggregate and has the deepest PMF curve among the different aggregate

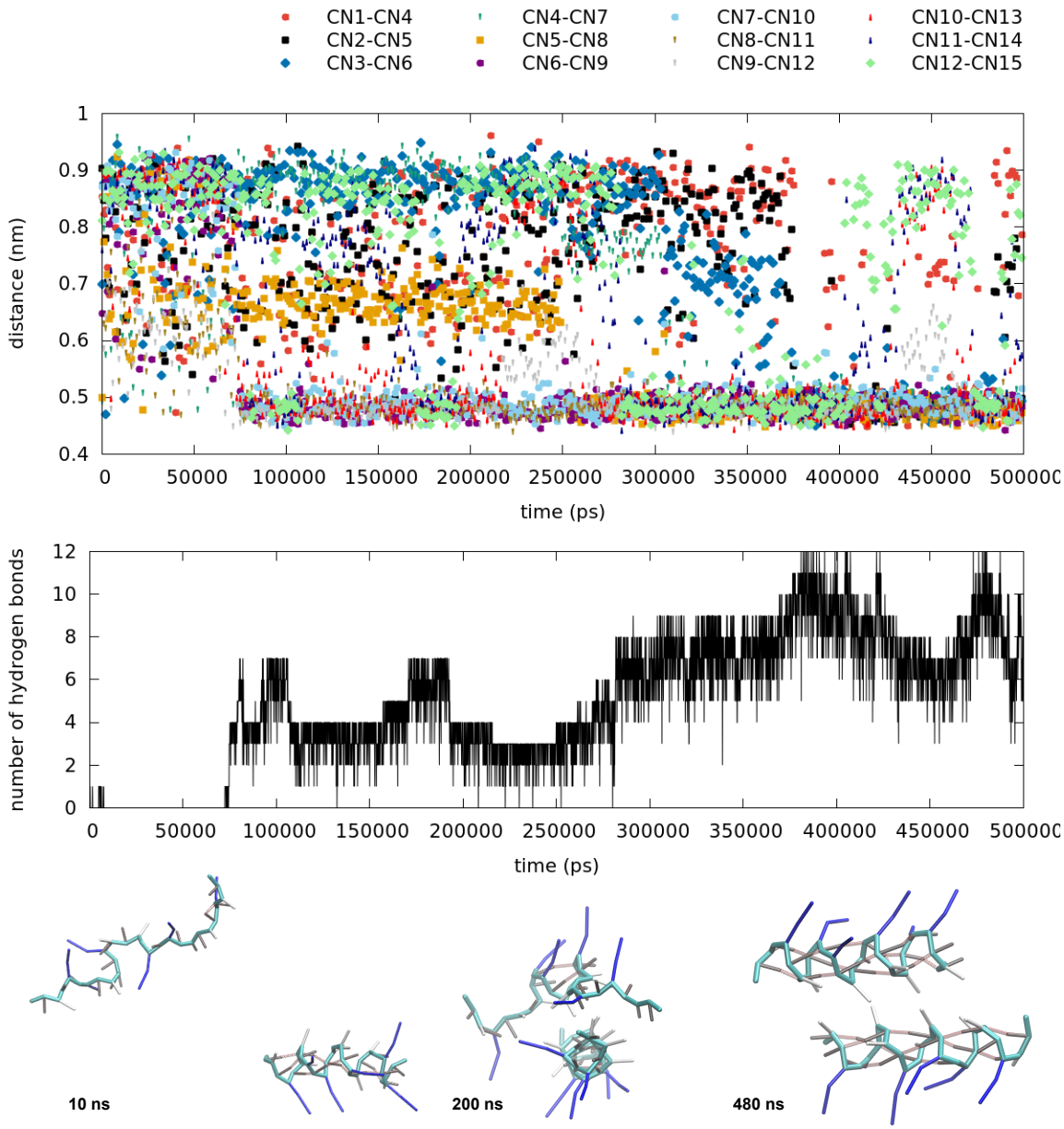

Figure 7.4: Hydrogen bonding distance $CN_i - CN_{i+3}$ versus time (top) of a single LK model, folded to an α-helix in the presence of molecular interface (another helix) in implicit solvent and while PME electrostatics has been taken into account. Number of hydrogen bonds versus time (bottom) also shows this transition to helix. E_w has been set to 18.5 kJ/mol, and E_{hb} is 17 kJ/mol and E_L 8.5 kJ/mol.

sizes. [1, 25] We have investigated the stability of tetramer by a free simulation of preassembled tetramer in implicit solvent for 100 ns. Resulting distance-angle

correlations have been plotted in Figure 7.5. Except folding-unfolding events for end residues, tetramer is stable and results in a similar correlation plots to atomistic tetramer.

Figure 7.6 shows PMF of separating one LK from a tetramer aggregate. In the aggregation mechanism, the interplay between the electrostatic repulsion of KNs and the attraction of the hydrophobic side chains define the stability of aggregates. Our model failed to mimic the atomistic PMF curve of separating peptide from tetramer, suggesting that leucine attraction is not strong enough to overcome this electrostatic repulsion in a tetramer aggregate.

Now that we have come up with this coarse-graining techniques and developed a two-state CG model that successfully mimics the bulk water and interface behaviour, we can extend it to other peptides that have a built-in secondary amphiphilicity upon forming α-helix or β-sheet. In other words the methodology is transferable to other amphiphilic peptides. In principle there is nothing specific to the LK molecule in this methodology and the scheme should be applicable for other peptides. One has to find the appropriate atomistic reference systems and parameterize bonded and non-bonded parameters separately. To correctly capture the conformational transition, nonbonded interactions have to be parameterized through the similar loops mentioned in each section.

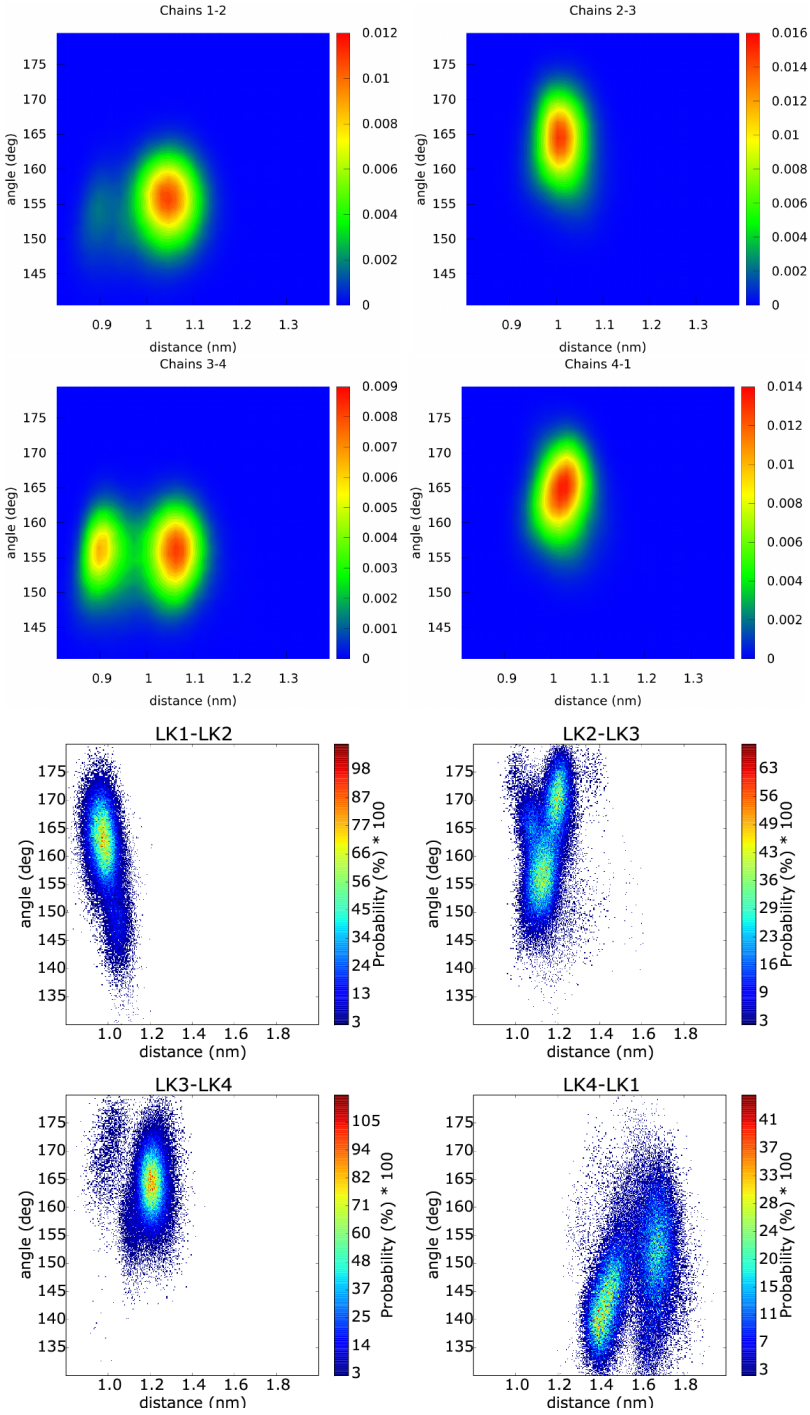

Figure 7.5: Distance versus angle correlation plots for LK dimers in a tetramer aggregate in atomistic (top) and CG (bottom) simulations. E_L 8.5 has been used together with E_{hb} 17 with preassembled helices in implicit solvent.

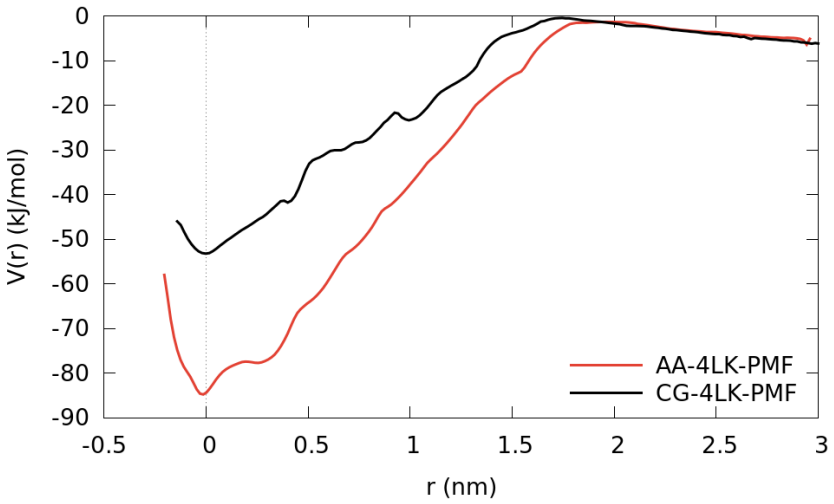

Figure 7.6: PMFs for separating LK from dimer in AA and CG model.

Direction depended hydrogen bonding potential using virtual representation of hydrogen has been used before in different CG models, and can be applied to CG models which have backbone hydrogen bonding. Hamiltonian replica exchange method has been used for a CG model in this study, to smooth the energy barrier between the folded helix and extended structure with enhanced sampling of conformational space. PMF of pulling two ends of helix can confirm the HREX results for E_{hb}. Coarse-grained interface has been constructed using density profiles of water slap and hydrophobic, hydrophilic and backbone residues in atomistic reference, and the adsorption strength of peptide to the interface leads us to tune the E_w. Transition from the random-coil conformation to α-helix at the interface would support the thermodynamic based approach used to obtain E_w. To mimic the behaviour of peptide in the presence of molecular interfaces (aggregates), radial distribution function of hydrophobic residues and potential of mean force for separation can be used to tune the shape and strength of hydrophobic potential.

REFERENCES

[1] Cahit Dalgicdir. Environment Induced Conformational Transition in Peptides. *PhD thesis, Graduate school of science and engineering, Koc University*, 2014.

[2] Emiliano Brini, Elena A Algaer, Pritam Ganguly, Chunli Li, Francisco Rodríguez-Ropero, and Nico FA van der Vegt. Systematic coarse-graining methods for soft matter simulations–a review. *Soft Matter*, 9(7):2108–2119, 2013.

[3] WG Noid. Perspective: coarse-grained models for biomolecular systems. *The Journal of chemical physics*, 139(9):090901, 2013.

[4] Tristan Bereau and Markus Deserno. Generic coarse-grained model for protein folding and aggregation. *The Journal of chemical physics*, 130(23):235106, 2009.

[5] E Lindahl, AR Hess, R v Buuren, E Apol, PJ Meulenhoff, DP Tieleman, ALTM Sijbers, KA Feenstra, R v Drunen, and HJC Berendsen. Gromacs user manual, version 4.5. 6., 2010. *There is no corresponding record for this reference*, 2010.

[6] Martin B Ulmschneider, Jacques PF Doux, J Antoinette Killian, Jeremy C Smith, and Jakob P Ulmschneider. Mechanism and kinetics of peptide partitioning into membranes from all-atom simulations of thermostable peptides. *Journal of the American Chemical Society*, 132(10):3452–3460, 2010.

[7] William F DeGrado, Holly Gratkowski, and James D Lear. How do helix–helix interactions help determine the folds of membrane proteins? perspectives from the study of homo-oligomeric helical bundles. *Protein science*, 12(4):647–665, 2003.

[8] Magnus Gustafsson, Guy Vandenbussche, Tore Curstedt, Jean-Marie Ruysschaert, and Jan Johansson. The 21-residue surfactant peptide (lysleu 4) 4 lys (kl 4) is

a transmembrane α-helix with a mixed nonpolar/polar surface. *FEBS letters*, 384(2):185–188, 1996.

[9] Dominik Fritz, Konstantin Koschke, Vagelis A Harmandaris, Nico FA van der Vegt, and Kurt Kremer. Multiscale modeling of soft matter: scaling of dynamics. *Physical Chemistry Chemical Physics*, 13(22):10412–10420, 2011.

[10] Parimal Kar and Michael Feig. Chapter five-recent advances in transferable coarse-grained modeling of proteins. *Advances in protein chemistry and structural biology*, 96:143–180, 2014.

[11] Jia-Wei Shen, Chunli Li, Nico FA van der Vegt, and Christine Peter. Transferability of coarse grained potentials: Implicit solvent models for hydrated ions. *Journal of Chemical Theory and Computation*, 7(6):1916–1927, 2011.

[12] Ozge Engin, Alessandra Villa, Christine Peter, and Mehmet Sayar. A challenge for peptide coarse graining: Transferability of fragment-based models. *Macromolecular Theory and Simulations*, 20(7):451–465, 2011.

[13] Dirk Reith, Mathias Pütz, and Florian Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of computational chemistry*, 24(13):1624–1636, 2003.

[14] Alexander P Lyubartsev and Aatto Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Physical Review E*, 52(4):3730, 1995.

[15] Cahit Dalgicdir, Ozge Sensoy, Christine Peter, and Mehmet Sayar. A transferable coarse-grained model for diphenylalanine: How to represent an environment driven conformational transition. *The Journal of chemical physics*, 139(23):234115, 2013.

[16] Jane R Allison, Sereina Riniker, and Wilfred F van Gunsteren. Coarse-grained models for the solvents dimethyl sulfoxide, chloroform, and methanol. *The Journal of chemical physics*, 136(5):054505, 2012.

[17] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and Wilfred F Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6. *Journal of computational chemistry*, 25(13):1656–1676, 2004.

[18] W Tschöp, Kurt Kremer, J Batoulis, Ta Bürger, and O Hahn. Simulation of polymer melts. i. coarse-graining procedure for polycarbonates. *Acta Polymerica*, 49(2-3):61–74, 1998.

[19] VA Harmandaris, NP Adhikari, Nico FA van der Vegt, and Kurt Kremer. Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations. *Macromolecules*, 39(19):6708–6719, 2006.

[20] Teemu Murtola, Emma Falck, Michael Patra, Mikko Karttunen, and Ilpo Vattulainen. Coarse-grained model for phospholipid/cholesterol bilayer. *The Journal of chemical physics*, 121(18):9156–9165, 2004.

[21] Furio Ercolessi and James B Adams. Interatomic potentials from first-principles calculations: the force-matching method. *EPL (Europhysics Letters)*, 26(8):583, 1994.

[22] Wataru Shinoda, Russell DeVane, and Michael L Klein. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation*, 33(1-2):27–36, 2007.

[23] Berk Hess, Christian Holm, and Nico van der Vegt. Osmotic coefficients of atomistic nacl (aq) force fields. *The Journal of chemical physics*, 124(16):164509, 2006.

[24] Sabine Castano, Bernard Desbat, Michel Laguerre, and Jean Dufourcq. Structure, orientation and affinity for interfaces and lipids of ideally amphipathic lytic l i k j (i= 2j) peptides. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1416(1):176–194, 1999.

[25] WF DeGrado and JD Lear. Induction of peptide conformation at apolar water interfaces. 1. a study with model peptides of defined hydrophobic periodicity. *Journal of the American Chemical Society*, 107(25):7684–7689, 1985.

[26] Victor Rühle and Christoph Junghans. Hybrid approaches to coarse-graining using the votca package: Liquid hexane. *Macromolecular Theory and Simulations*, 20(7):472–477, 2011.

[27] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David van der Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, page btt055, 2013.

[28] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[29] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

[30] Glenn M Torrie and John P Valleau. Monte carlo free energy estimates using non-boltzmann sampling: application to the sub-critical lennard-jones fluid. *Chemical Physics Letters*, 28(4):578–581, 1974.

[31] John G Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 1935.

[32] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry*, 16(11):1339–1350, 1995.

[33] Jochen S Hub, Bert L De Groot, and David Van Der Spoel. g_wham? a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation*, 6(12):3713–3720, 2010.

[34] Giovanni Bussi. Hamiltonian replica exchange in gromacs: a flexible implementation. *Molecular Physics*, 112(3-4):379–384, 2014.

[35] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.

[36] Massimiliano Bonomi, Davide Branduardi, Giovanni Bussi, Carlo Camilloni, Davide Provasi, Paolo Raiteri, Davide Donadio, Fabrizio Marinelli, Fabio Pietrucci, Ricardo A Broglia, et al. Plumed: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, 180(10):1961–1972, 2009.

[37] Charles D Schwieters and G Marius Clore. The vmd-xplor visualization package for nmr structure refinement. *Journal of Magnetic Resonance*, 149(2):239–244, 2001.

[38] Jong Cheol Jo and Byeong Cheol Kim. Determination of proper time step for molecular dynamics simulation. *Bulletin of the Korean Chemical Society*, 21(4):419–424, 2000.

[39] Anne Voegler Smith and Carol K Hall. α-helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins: Structure, Function, and Bioinformatics*, 44(3):344–360, 2001.

[40] DK Klimov, MR Betancourt, and D Thirumalai. Virtual atom representation of hydrogen bonds in minimal off-lattice models of α helices: effect on stability, cooperativity and kinetics. *Folding and Design*, 3(6):481–496, 1998.

[41] Eng-Hui Yap, Nicolas Lux Fawzi, and Teresa Head-Gordon. A coarse-grained α-carbon protein model with anisotropic hydrogen-bonding. *Proteins: Structure, Function, and Bioinformatics*, 70(3):626–638, 2008.

[42] Anders Irbäck, Fredrik Sjunnesson, and Stefan Wallin. Three-helix-bundle protein in a ramachandran model. *Proceedings of the National Academy of Sciences*, 97(25):13614–13618, 2000.